University of Alberta

Economic modeling of oil and gas exploration activities in Alberta

by

Michael Habteyonas ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment

of the requirements for the degree of Doctor of Philosophy

in

Agricultural and Resource Economics

Department of Rural Economy
Edmonton, Alberta
2008

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-46326-0
Our file Notre référence
ISBN: 978-0-494-46326-0

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The energy sector is the key driver of Alberta's economy. However, there are externalities associated with the sector. A number of ecological and environmental concerns related to landscape changes and wildlife habitat destruction have been raised in recent years. Improved understanding of oil and gas exploration activities will be useful in assessing the potential impact of energy sector activities on the environment as well as the impact of policy options on exploration behavior. The thesis is organized in three parts.

The first part develops an econometric model of exploration activities by addressing the following questions:

- What mix of economic and geological factors affect exploration activity in Alberta?
- How does the energy sector respond to uncertainty associated with exploration?
- How does exploration activity vary spatially across regions in the province?

A spatial exploration effort model is estimated using drilling density as the dependent variable and economic and geological factors as explanatory variables. Regional models within Alberta are also estimated to capture the problem of spatial heterogeneity. Significant differences were observed among the coefficients estimated for Alberta and the three regions within Alberta.

The second part of the thesis forecasts future drilling activity on the landscape of Alberta up to the year 2020. The model developed in the first part is applied to forecast spatial exploration. The results show that forecasts made using the spatial model perform better than the non-spatial model for the three regions. Sensitivity of drilling forecasts to price changes were examined assuming 10, 20, 50 and 100 percent increase in prices.

The main objective of the third part of the thesis was to test the hypothesis that the energy sector anticipates new environmental regulations and increases exploration activities prior to the regulations being implemented. Three models, a multivariate regression model, the difference in difference method, and propensity score matching methods were used to test the hypothesis. The results indicate that on average drilling density in caribou habitat was higher than in non-caribou habitat. Furthermore, the results suggest that listing caribou as a threatened species has not resulted in reductions in exploration effort in caribou regions and may have generated additional effort in the region. The latter finding can be interpreted as evidence that the sector anticipates potential restrictions on land use.

Acknowledgments

I thank God for giving me the strength and health to start and finish my study.

In the first place, I would like to give special thanks to my supervisor Vic Admowicz for his continuous advice and guidance from the early stage of this research. I am indebted to him more than he knows. I gratefully acknowledge my co-supervisors Marian Weber and Grant Hauer for their advice and crucial contribution to this thesis. Special thanks to Marian for reviewing the whole paper. I would also like to thank Sean Cash for being willing to be in my supervisory committee member. I have benefited from his advice and guidance. I thank Peter Boxall for sitting in my candidacy exam and for his valuable comments. Special thanks to Stan Boutin and Brady Deaton for their willing to be my external examiners. I am grateful to Stan for his time and comments on the last chapter of the thesis too. I would like to thank the Department of Rural Economy, especially Ellen Goddard, for providing me with additional funding to cover my international differential fees and expenses related to attending conferences. Thanks to Robin McClelland for all the administrative help. I also thank the Faculty of Graduate Studies and Research for funding support.

Many people and organizations have made this thesis a success by providing funding, advice, data, software programs and routines. Special thanks to the Sustainable Forest Management Network (SFMN) and the BEEST project for providing me with funding. I would like to thank members of the BEEST project, Robert Jagodzinski, Xianli Wang and Steve Cumming for providing me data and technical advice related to GIS. Special

thanks to GeoSCOUT™ for allowing us to use their oil and gas database. Thanks to the Alberta Department of Energy and the Alberta Energy and Utilities Board (EUB) for providing us with data on prices and reserves. I would also like to thank Paul Elhorst for allowing me to use his Matlab routines on spatial panel data estimation and Thomas Love for the propensity score matching algorithm.

Special personal thanks to my friends Jing Zhang, Tsion Demeke, Pastor Mengesteab Mengistu, and members of the Eritrean Evangelical Church in Edmonton for their support and encouragement. Last but not least, to my beloved family in Eritrea, especially to my dear sisters Weyni, Welesh, and Lemlem for their support and encouragement. Finally, I would like to thank everybody who was important to the successful realization of the thesis, as well as expressing my apology that I could not mention personally one by one.

Table of Contents

Abstract

Acknowledgments

List of Tables

List of Figures

Chapter 1 Introduction 1

1.1 Background 1

1.2 Research problem and objectives of the study 4
- 1.2.1 Econometric model of exploration effort 5
- 1.2.2 Forecasting exploration effort 6
- 1.2.3 Energy sector anticipation of environmental regulation 6

1.3 Contribution of the study 7

1.4 Organization of the thesis 8

Chapter 2 Theory, Literature Review, and Empirical Model 9

2.1 Economic versus geologic models of exploration and development 9

2.2 Ecological studies 11

2.3 Economic theory 12
- 2.3.1 Optimal exploration of oil and gas 13
- 2.3.2 Exploration with uncertainty and learning 16
- 2.3.3 Spatial exploration and regional models 20

2.4 Empirical Model 22
- 2.4.1 Spatial econometric issues 25
- 2.4.2 Spatial specification of the model 27

Chapter 3 Data and Descriptive Analysis 31

3.1 Data sources and description 31

3.2 Map of the study area 36

3.3 Descriptive statistics 38

3.4 Exploratory data analysis 38

3.5 Testing for spatial autocorrelation 43

Chapter 4 Model Results 47

4.1 Model estimation 47

4.2 Spatial lag versus spatial error models 50

4.3 Regional models 53

4.4 Discussion of results 61

4.5 Summary 67

Chapter 5 Forecasting oil and gas exploration activities 69

5.1 Introduction 69

5.2 The forecasting model and data 71
5.2.1 The process of forecasting 71
5.2.2 Projected data for explanatory variables 73

5.3 Model validation 76
5.3.1 The process of validation 76
5.3.2 Coefficient estimates for the validation model 77
5.3.3 Comparing ex post vs. actual forecasts 81

5.4 Model results and forecasts 84

5.5 An application of the forecasting model 94

5.6 Summary and discussion of the results 99

Chapter 6 Analysis of the energy sector's anticipation of environmental regulations **101**

6.1 Introduction 101

6.2 Description of the hypothesis test 106

6.3 Caribou ranges and study area 107

6.4 Data and descriptive analysis of exploration in caribou habitat 111
6.4.1 Data 111
6.4.2 Descriptive analysis of exploration in caribou habitat 111

6.5 Multivariate regression 116
6.5.1 Multivariate model 116
6.5.2 Results 117

6.6 The difference in difference method 124
6.6.1 Introduction 124
6.6.2 Model and methods 125
6.6.3 Results for the difference in difference model 128

6.7 Propensity score matching 131
6.7.1 Introduction 131
6.7.2 Methodology and data 131
6.7.3 Propensity Score Matching Results – Part I 136
6.7.4 Propensity Score Matching Results – Part II 149

6.8 Summary of results and hypothesis tests 151

6.9 Management implications 154

Chapter 7 Conclusions **157**

References **162**

Appendixes **174**

Appendix 1.1 Alberta Township Survey *174*

Appendix 2.1 Derivation of optimal exploration equation *176*

Appendix 2.2 Matlab routines for estimating spatial panel data models *183*

Appendix 6.1 Average drilling densities by herd name *186*

Appendix 6.2 Map based on Petroleum Services Association of Canada (PSAC) *189*

Appendix 6.3 Results for the herd by herd multivariate regression model *190*

List of Tables

Table 3.1 Data required to estimate the empirical model and description of the variables 31
Table 3.2 Descriptive statistics of variables for the three regions in Alberta 39
Table 3.3 Moran's I test for years 1985, 1992, 2000 and 2004. 44
Table 4.1 Results for aspatial and spatial models of drilling density in Alberta 48
Table 4.2 Likelihood Ratio (LR) test for spatial versus aspatial models 50
Table 4.3 Results for the spatial lag models of drilling density in Alberta 51
Table 4.4 Results for the spatial error models of drilling density in Alberta 52
Table 4.5 Results for regional spatial lag models 54
Table 4.6 Results of Likelihood Ratio (LR) test for the spatial heterogeneity hypothesis: Spatial lag model 55
Table 4.7 Results for regional spatial error models 59
Table 4.8 Results of Likelihood Ratio (LR) test [illegible] the spatial heterogeneity hypothesis: Spatial error model 60
Table 4.9 Summary of price elasticities of drilling effort for the spatial lag and spatial error models 63
Table 5.1 Results of the success rate equation: model validation sample (1980 – 2000). 79
Table 5.2 Results of the drilling effort equation: model validation sample (1980 – 2000) 80
Table 5.3 Comparison of actual vs. ex post forecasts using statistical indicators 83
Table 5.4 Results of the drilling effort equation for the three regions: (1980 – 2004) 85
Table 6.1 Results for the multivariate regression model of caribou habitat: Model 1 118
Table 6.2 Summary of results for the multivariate regression model estimated using dummy variables for each herd: Model 2 herd by herd model. 120
Table 6.3 Multivariate regression results for each region: Model 3 - PSAC regional model 122
Table 6.4 Results based on the difference in difference model for the two scenarios:... 130
Table 6.5 Results of the logit model used constructing the propensity score index 137
Table 6.6 Statistical tests of mean difference of the covariates for the matched groups 142
Table 6.7 Mean difference of well densities for the matched groups: all herds 147

Table 6.8 Mean difference of well densities for the matched groups: Cold Lake and Salve Lake herds 149
Table 6.9 Results for mean difference of covariates and well densities: excluding forest cover and muskeg. 150

List of Figures

Figure 1.1 Trend of oil and gas exploration and development wells in Alberta 2
Figure 3.1 Alberta regional boundary map 37
Figure 3.2 Summary of wells drilled between 1980 and 2004. 40
Figure 3.3 Trend of regional drilling activities in Alberta 41
Figure 3.4 Wells drilled through time for a sample of townships with intensive drilling 42
Figure 3.6 Scatter plot of Moran's I for the year 1982 44
Figure 3.7 Local indicator of spatial autocorrelation (LISA) map for the year 2000 46
Figure 4.1 Temporal price elasticity of drilling: Alberta 64
Figure 4.2 Temporal price elasticity of drilling: Northern region 64
Figure 4.3 Temporal price elasticity of drilling: Plains region 65
Figure 4.4 Temporal price elasticity of drilling: Foothills region 65
Figure 5.1 Forecasts of average price of oil at Alberta wellhead (Source EUB 2005) 75
Figure 5.2 Forecasts of average price of natural gas at plant gate (Source EUB 2005) ... 75
Figure 5.3 Northern region: actual versus ex post forecasts 81
Figure 5.4 Plains region: actual versus ex post forecasts 82
Figure 5.5 Foothills region: actual versus ex post forecasts 82
Figure 5.6 Summary of forecasts for the northern region 87
Figure 5.7 Summary of forecasts for the plains region 87
Figure 5.8 Summary of forecasts for the foothills region 88
Figure 5.9 Map of actual wells for the year 2004. 89
Figure 5.10 Map of forecasted wells for the year 2005. 90
Figure 5.11 Map of forecasted wells for the year 2010 91
Figure 5.12 Map of forecasted wells for the year 2015 92

Figure 5.13 Map of forecasted wells for the year 2020. 93
Figure 5.14 Map of forecasted WTSP bird counts for the year 2010. 96
Figure 5.15 Map of forecasted WTSP bird counts for the year 2015. 97
Figure 5.16 Map of forecasted WTSP bird counts for the year 2020. 98
Figure 6.1 Woodland caribou range boundaries in Alberta 107
Figure 6.2 Histogram of number of townships for each caribou herd 109
Figure 6.3 Map of study area 110
Figure 6.4 Trend of average well densities in caribou and non-caribou habitat 112
Figure 6.5 Average well densities in Little Smoky, Salve Lake and non-caribou Habitat 113
Figure 6.6 Average well densities in Cold Lake, WSAR, ESAR, and non-caribou Habitat 114
Figure 6.7 Average density of drilling aggregated into four regions 116
Figure 6.8 Graphical explanation of the difference in difference (DID) method 127
Figure 6.9 Map of matched caribou and non-caribou habitat: all study areas 139
Figure 6.10 Map of matched caribou and non-caribou habitat: A La Peche Herd 139
Figure 6.11 Map of matched caribou and non-caribou habitat: Cold Lake Herd 140
Figure 6.12 Map of matched caribou and non-caribou habitat: Red Earth 1441
Figure 6.13 Average drilling densities for the matched caribou habitat and non-caribou habitat: All study area 145
Figure 6.14 Average drilling densities excluding Caribou Mountain and Bistcho herds 144
Figure 6.15 Average drilling densities excluding Caribou Mountain , Bistcho, and Cold Lake herds 145
Figure 6.16 Average well densities for the matched caribou and non –caribou habitat: A La Peche 145
Figure 6.17 Average drilling densities for the matched caribou and non-caribou habitat: Red Earth 146
Figure 6.18 Average drilling densities for the matched caribou and non-caribou habitat: Cold Lake 146
Figure 6.19 Hectares of land leased and licensed in Alberta 155

Chapter 1 Introduction

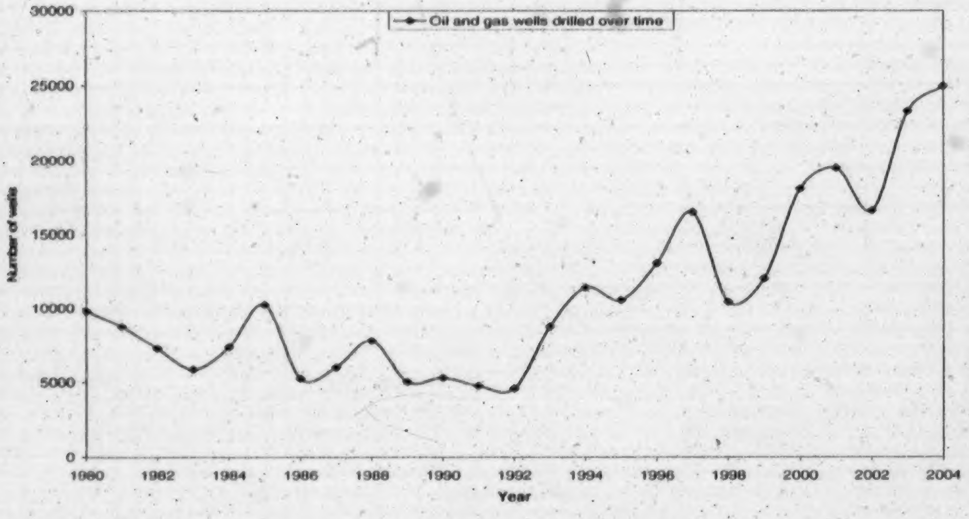
1.1 Background

The energy sector plays a leading role in Alberta's economy. Energy related royalty revenues account for about one-third (about $14.7 billion in 2005-06) of the total revenue collected by the province of Alberta (DOE 2006). These revenues are critical to the delivery of public programs such as health-care and education. Energy also accounts for over half the value of the province's $66 billion in total exports and about one-quarter of the total $170 billion in Gross Domestic Product (DOE, 2005). About 17 percent of the workers in Alberta are employed directly or indirectly in the province's energy sector (DOE 2005). In Alberta, the Crown owns 81 per cent of the province's mineral rights and the remaining 19 per cent are 'freehold' mineral rights owned by the federal government (on behalf of First Nations or in National Parks) and by individuals and companies (DOE, 2005). The Department of Energy administers the mineral rights in the form of licenses or leases tendered through a competitive sealed bid auction (DOE, 2005). The province holds an average of 24 land auctions each year and issues approximately 8,000 petroleum and natural gas agreements per year (DOE, 2005).

Once an oil and gas lease is issued, the process of oil and gas development takes places in a series of stages. Industries start to explore for oil and gas using preliminary geophysical exploration methods. At a preliminary stage, seismic line prospecting is the most common method used for locating subsurface structures that may contain oil or/and gas.

Favorable seismic results will then lead to further exploration drilling[1]. Upon completion of drilling, the well is cased and tested to obtain information about the rock formation and production of fluids. If oil or gas is not discovered in commercial quantities, the well is considered dry.

Figure 1.1 Trend of oil and gas exploration and development wells in Alberta



Data source: Data extracted and summarized from GeoSCOUT™

Figure 1.1 shows the trend of oil and gas drilling activity since 1980. Drilling activity has been increasing since 1992 with some downward cycles in 1998 and 2002. Part of the reason for the upward trend of drilling activity after 1992 could be the introduction of the third tier exploratory royalty holiday[2]. This policy was introduced by the Department of Energy to encourage the discovery of new oil reservoirs. The introduction of new 3D seismic surveys in the 1990s could also be another factor for the upward trend of drilling activities[3]. Given large deposits of oil sands and coal bed methane in Alberta, exploration and development wells are expected to increase tremendously.

---

[1] Seismic lines are a long linear corridor cleared using a bulldozer in which series of holes are drilled along the corridor and dynamite charges are placed. The dynamite charges are sequentially exploded and the reflected sound waves are recorded at the surface using portable recording equipment. Such seismic (or geophysical) exploration is used to identify and map potential oil and gas deposits prior to drilling (Schneider 2002).

While the energy sector is the primary driver of Alberta's economy, there are externalities associated with the sector. Oil and gas exploration requires infrastructure that creates linear features on the landscape. Clearing of trees for the construction of seismic lines is associated with progressive loss and fragmentation of wildlife habitat. Construction of well sites and facilities, access roads, and pipelines are linked to increased access to humans and other predators of wildlife. An example of the ecological impact of energy sector activities is on caribou habitat. Woodland caribou in Alberta is one of the species likely to be extirpated from regions with intensive fragmentation of the landscape (Dzus 2001). Previous studies have shown that caribou declines across Alberta have been correlated with the level of industrial development within their ranges (Dzus 2001). Schneider et al (2003) comment that the main factor for the decline in caribou is not the impact of individual wells but the cumulative impact of all wells and other infrastructure related to the construction of wells. In 2005 alone, 21,599 new wells were drilled in Alberta and the total extensive pipeline infrastructure to 2003 amounted to 332,464 kilometers (CAAP, 2005). Several recent studies have addressed the impact of exploration on caribou population (Schneider and Dyer 2006). A government-led study concluded that woodland caribou will continue to decline unless limits to development and aggressive restorations of existing disturbances are implemented (East Side Range Planning Team 2005). An industry-funded modeling study within the oil sands region determined that, due to projected industrial development, available caribou habitat will decline from 43 percent of the landscape to 6 percent over the next 20 years (Schneider et al 2003). A third study by Weclaw and Hudson (2004) concluded that caribou will be extirpated from northern Alberta in less than 40 years if linear densities exceed 1.2 km/km^2. Understanding the behavior of the energy sector drilling activities and its future impact is essential in order to reduce the negative impacts of drilling on caribou ranges and other conservation objectives and to make an informed policy decisions.

---

[2] The third tier exploratory royalty holiday is a permanent policy to encourage the discovery of new oil reservoirs. It is applicable to exploratory oil wells started on or after October 1, 1992 (Department of Energy Information Letter 93 -8).

[3] The above wells do not include oil sands and coal bed methane.

1.2 Research problem and objectives of the study

Empirical research on how the energy sector decides where and when to explore for oil and gas resources is limited. Most analysis of exploration effort for a particular basin is based on aggregated regional models. The first research problem of this study is to identify the factors that affect oil and gas exploration effort and to develop and estimate a spatial and temporal model of exploration activities within the province of Alberta. The model will then be used to forecast drilling activities and to analyze the energy sector's anticipation of environmental regulations in wildlife areas with a special focus on woodland caribou habitat.

1.2.1 Econometric model of exploration effort

The first objective is to develop an econometric model of oil and gas exploration activities on the landscape of Alberta. The model is developed based on economic theory and review of the literature. Historical data from Alberta's oil and gas drilling activities are used to estimate the model and empirically determine the spatial and temporal factors that affect exploration. A common practice in modeling exploration and development of non-renewable resources is to develop an inter-temporal model of profit maximization (Pindyck 1978). The present study builds on the previous studies by developing an exploration model using spatial econometric procedures that take into account spatial interactions as well as inter-temporal dimensions in the conventional multivariate regression model. Spatial dependence is incorporated for theoretical and empirical reasons.

Regional models within the province of Alberta are estimated to capture differences in geological processes. The province is divided into three regions based on geological and geographic differences. These regions are the Plains, Northern, and Foothills regions. Exploration behavior at a regional level is obscured when aggregated regional data are used (e.g. Attanasi 1979, Siegel 1985, Cairns 1990, Kunce et al 2004). The present study utilizes spatially disaggregated data at a resolution of 10 km by 10km grids of land within the province of Alberta[4]. The main advantage of using disaggregated spatial data is to understand firms' exploration behavior at a specific site. The other advantage is that many environmental impacts of exploration are spatially dependent. Geological and economic variables that capture the underlying framework of the energy sector behavior are included in the model. These variables include the *cumulative number of wells, lagged success rate, wellhead prices of oil and gas, capacity utilization rate and technological changes.*

---

[4] The 10km by 10km grids of land are commonly called 'Townships'. Refer Appendix 1.1 for a detailed description of the Alberta township system.

1.2.2 Forecasting exploration effort

The second objective of the study is to develop spatial and temporal forecasts of oil and gas exploration activities based on the estimated econometric model. The forecasting process is performed at a township level using scenarios for exogenous variables such as the price of oil and gas[5]. The forecasts can be used in the future to simulate possible trade offs between economic development opportunities and biodiversity conservation in the boreal plains region.

1.2.3 Energy sector anticipation of environmental regulation

The third objective of the study is to analyze how anticipation of new regulatory announcements related to habitat protection would affect energy sector behavior. This is implemented by testing the hypothesis that the energy sector would anticipate environmental restrictions and increases exploration activities in wildlife areas where land use restrictions may arise. This is related to the issue of fear of regulation or anticipation of new regulations that protect wildlife. For example Lueck and Michael (2003) found evidence that landowners preemptively destroyed habitat for the endangered red-cockaded woodpeckers (RCWs) in the forests of North Carolina in order to avoid potential land use regulations prescribed under the Endangered Species Act (ESA). In the present study, three statistical methods are used to compare intensity of drilling activities in caribou habitat and non-caribou habitat. These are multivariate regression methods, the difference in difference method, and propensity score matching method.

---

[5] This work is an integral part of the Boreal Ecology and Economics Synthesis Team (BEEST) project. The aim of the BEEST project is to improve our understanding of the spatial economic behavior of different sectors.

1.3 Contribution of the study

The present study contributes to the literature on exploration of oil and gas in many ways. The first part investigates the spatial dimensions of the behavior of the energy sector on the landscape. The effects of historical patterns of exploration, as well as economic and geological factors that affect exploration are empirically determined using spatial econometric techniques. Furthermore, hypotheses related to learning and clustering of oil and gas activities, and the depletion effects of resources are tested. The second part of the thesis uses various forecasting approaches to forecast spatial oil and gas drilling to the year 2020. Results from the forecasting model can be used to simulate changes in ecological indicators such as bird abundance models. The forecasting results are also useful for land use planning and management issues related to the conservation of ecological resources on the landscape of Alberta. Finally, the study assesses the extent to which the energy sector anticipates and responds to regulations that may limit access prior to the implementation of the regulations.

1.4 Organization of the thesis

The thesis is organized as follows: Chapter one contains an introduction to the oil and gas exploration activities in Alberta. The research problem, the main objectives, and contribution of the study are discussed in this chapter. The first part of chapter two reviews the theory and literature on non-renewable resources in general and oil and gas exploration in particular. The second part of the chapter develops the empirical model. Chapter three describes the study area and the data. Exploratory analyses of the data and tests on spatial autocorrelation are outlined in this chapter. Model results based on different specifications of the spatial model are presented in chapter 4. The first section compares spatial and non-spatial models. Within the family of spatial models the next section compares the spatial lag and spatial error models. Section three presents results based on regional models and the last sections summarize and conclude the chapter. Chapter five examines forecasting oil and gas drilling activities. The model developed in chapter two and results obtained from chapter 4 are used to forecast future drilling activities. Chapter 6 is an application of the oil and gas exploration model to test the hypothesis of energy sector's anticipation of environmental regulations. This chapter deals with how anticipation of environmental regulations would affect oil and gas exploration activities in woodland caribou habitat. The final sections of the thesis are devoted to conclusions, references, and appendices.

Chapter 2 Theory, Literature Review, and Empirical Model

The purpose of this chapter is to review the literature and to develop the specification of the empirical model to be used in explaining and forecasting energy sector exploration activity. The first and the second sections discuss different approaches to non-renewable resources studies in economics, geology and ecological science literature. A comparison between these approaches is discussed and the present study's approach is outlined. Section three explains the economic theory behind oil and gas exploration and discusses different issues raised in the economic literature. For instance, uncertainty in the process of exploration and how firms address uncertainty using learning models is discussed. Spatial exploration and regional models of exploration are also discussed in this section. The empirical model is developed in section four.

2.1 Economic versus geologic models of exploration and development

The study of non-renewable resources, specifically the economics of exploration, is a vast topic. Devarajan and Fisher (1981) conducted a detailed survey of the literature on non-renewable resources. The authors reviewed studies on the economics of exhaustible resources since the seminal work of Hotelling's (1931) paper up to 1980s. Cairns (1990, 1994) reviewed microeconomic supply models linked to the exploration of a small region or a play[6]. Dahl and Duggan (1998) have also done a survey of U.S. models of oil and gas exploration with a special focus on price elasticities of exploration effort based on economic exploration models. Various models have been proposed to estimate and forecast the time path of resource exploration and extraction. Some of these models have been dominated by geological considerations and others by economic aspects.

---

[6] A play is an area of concentrated similar exploration activities within a sedimentary basin (Cairns 1990).

Walls (1992) compares geologic and econometric approaches to modeling the supply of oil and gas. The geologic models are classified into play analysis and discovery processes models (Walls 1992). Play analysis models are primarily used in relatively unexplored areas and rely on detailed geologic data. Discovery process models are generally used in well developed areas where information on exploration activity and oil and gas discovery sizes is readily available (Walls 1992). Econometric models, on the other hand, focus on the estimation of historical relationships between economic variables and drilling and reserve additions (Fisher 1964).

The best known geological approach is the Hubbert model (1956). This model is based on the proposition that forecasts of oil production can be obtained through the use of a logistic curve of cumulative petroleum discoveries and the exponential decline curve of yield per unit effort. Hubbert's work was criticized for its preoccupation with geological and geophysical phenomena and the exclusion of economic factors (Lynch 2002). Kaufmann and Cleveland (2001) and Moroney and Berg (1999) have established that the purely geologic and purely economic models both suffer from the common flaw of model misspecification and consequent prediction errors. While both types of models can yield unreliable predictions of production, due the exclusion of geologic variables, economic models may even produce unexpected results in which oil prices and production move in opposite directions (Kaufmann and Cleveland 2001; and Moroney and Berg 1999). The authors suggest that models of exploration and extraction that integrate geologic and economic factors may yield superior results than either the purely geologic or purely economic models. The present study develops an exploration model that incorporates both economic variables such as wellhead price of oil and gas and geologic variables such as *success rate* of wells and *cumulative drilling* activities that capture the geological state of the region.

2.2 Ecological studies

Understanding the spatial and temporal behavior of the upstream energy sector has key implications for land-use management and conservation of ecosystems. For the case of Alberta, Schneider (2002) and Schneider et al (2003) have discussed the ecological impacts of energy sector activities in Alberta's boreal forest. The main objective of these studies is to understand how oil and gas exploration and development contributes to the fragmentation of Alberta's forests and wildlife habitat. The authors showed that the area of forest cleared by seismic explorations is almost equal in size to harvest by the forest industry. For instance, on the Al-Pac Forest Management Agreement (FMA) area the rate of harvest was 16,000 ha/year by the forest industry and 11,000 ha/year by the petroleum industry (Schneider, 2002). Similarly Severson-Baker (2004) outlines ecological and environmental impacts associated with exploration and production in the northern part of Canada. The study qualitatively describes disturbance of land surfaces, damage to vegetation and soil, and ground water contamination due to energy sector development in the North.

The above mentioned ecological studies assume that exploration of oil and gas wells follow simple trends or projections suggested by the Hubbert model (Schneider et al 2003). These studies do not specify the underlying behavioral framework of the energy sector. For example, reserves of oil and gas tend to be clustered rather than distributed randomly. This leads to the discovery of new deposits closer to known deposits and leads to excess activity in the play (Cairns 1990). Petroleum industries engage in exploration and extraction activities to maximize the present value of their profits subject to reserve and economic constraints. Solving the optimization problem gives us an optimal time path for exploration and extraction under various price and policy regimes. Exploration or drilling effort is determined by a number of economic and geologic factors. These factors include distribution and extent of reserves, provision of infrastructure and capacity utilization, expected present value of oil and gas at a given time, technological change, and regulatory and institutional factors. In order to understand how the energy sector decides to explore for oil and gas, a detailed behavioral model that captures these factors needs to be developed.

2.3 Economic theory

In the economics of non-renewable resources, it is often assumed that perfectly competitive industries are involved in exploration and extraction activities and these industries maximize the discounted present value of future operating profits from the sale of a resource. The common practice is to build an aggregated model of inter-temporal profit maximization subject to the constraints of underlying reserves and technology

(Pindyck 1978, Peseran 1990). Some authors have identified problems of aggregating across fields and addressed the issue by constructing models of exhaustible resources at a field level (Livernois and Uhler 1987, Quyen 1991, Gaudet et al 2001). The theoretical model developed below is based on previous studies, but is tailored towards the spatial behavior of exploration.

2.3.1 Optimal exploration of oil and gas

The following assumptions are made to develop the theoretical model. First, perfectly competitive producers are assumed to maximize the discounted present value of profits from the sale of oil and gas. A single firm is used to represent the industry; hence the common pool problem is not considered. Second, the main purpose of exploration is to find new reserves and obtain information about potential reserves. Therefore the exploration program is based on information accumulated from the previous periods and neighboring townships. Third, oil and gas are treated jointly in the analysis rather than as separate resources because wells are classified as oil, gas, or dry only after the outcome of drilling is known.

The province of Alberta is considered to be an exploration region partitioned into N grids of land, where N is a positive integer. We assume that each grid is either empty or contains a certain size of potential deposits. In our case N is specified as the number of townships in Alberta[7]. The firm decides on the optimal exploration effort (w_{it}) and extraction of oil and gas (q_{it}) by maximizing the expected present value of profits subject to the constraints[8]. Formally, this can be written as follows:

---

[7] Description of the Alberta township system is given in Appendix 1.1.

$$\underset{w_{it}, q_{it}}{Max}.E\int_{0}^{T}\left\{\sum_{i=1}^{N}P_t q_{it}-\sum_{i=1}^{N}C^{1i}(w_{it},\Delta_{it})-\sum_{i=1}^{N}C^{2i}(q_{it},R_{it},\Lambda_{it})\right\}e^{-\delta t}dt \qquad 2.1$$

Subject to: $$\frac{dR_{it}}{dt}=d_{it}-q_{it} \text{ and } \quad q_{it}<R_{it} \qquad 2.2$$

$$d_{it}=f(w_{it},A_{it-1})^{9} \qquad 2.3$$

$$q_{it}\geq 0, w_{it}\geq 0, R_{it}\geq 0, A_{it}\geq 0 \qquad 2.4$$

where i refers to townships 1, 2, 3,…N; t indicates time period 1,2,3,…T, w_{it} is drilling effort at township i in period t, q_{it} refers to extraction of a resource at township i in period t, P_t represents well head price of a resource at time t, and δ is the discount rate. C^{1i} (.) and C^{2i} (.) show costs of exploration and extraction at township i respectively. R_{it} refers to proven reserves at township i in period t, d_{it} shows reserve additions, A_{it-1} refers to cumulative exploratory effort at township i in period t-1. Δ_{it} and Λ_{it} represent vectors of exogenous physical factors such as technology and *capacity utilization* related to exploration and extraction of oil and gas at township i in period t respectively.

---

[8] The expectation operator is used to capture the stochastic nature of the optimization problem. This could include uncertainties that include the risk of drilling a dry hole, economic uncertainties related to demand and future oil and gas prices, risk that a discovery will not be large enough to recover initial exploration costs, reserve size uncertainties, and so on.

[9] In equation 2.3 A_{it-1} refers to lag of cumulative exploratory effort. This specification is used based on the empirical results of Uhler (1976) and Pesaran (1990).

Equation 2.1 is the objective function where E refers to expectation operator and equations 2.2 to 2.4 are the constraints. Equation 2.2 shows that the change in reserves (dR/dt) is explained by the difference between reserve additions and extraction, and firms can not extract more than current reserves ($q_{it} < R_{it}$). Equation 2.3 explains that current reserve additions are determined by current exploratory effort (w_{it}) and past development represented by the lag of cumulative exploratory effort A_{it-1}. The optimal time path for the exploratory effect (w_{it}) equation can be obtained by manipulating the optimality conditions and solving the above equations.

Solving the first order conditions of the equations 2.1 to 2.4 yields[10]:

$$\frac{dw_i}{dt} = E\left(\frac{C^{1i}_{w_{it}}(\cdot)\left[(f_{(wA)_{it}} / f_{w_{it}}) \cdot f(w_{it}, A_{it-1}) - f_{A_{it}} + \delta\right] + C^{2i}_{R_{it}}(\cdot) f_{w_{it}}}{\left[C^{1i}_{(ww)_{it}} - C^{1i}_{w_{it}}(\cdot) f_{(ww)_{it}} / f_{w_{it}})\right]}\right) \qquad 2.5$$

Equation 2.5 shows the time path of exploratory effort. This equation implies that drilling effort is determined by a complex interaction between the expected cost of finding new reserves (expected marginal exploration costs, C^1_w), expected marginal extraction cost due to stock effects assuming R is not stochastic (C^2_R), expected marginal product of exploratory effort (f_w), expected reserve additions $f(w, A)$, and the rate of interest. Expected costs of exploration and extraction are in turn affected by the level of technology, initial level of reserves, and price of the resource. The expected reserve additions are determined by the level of previous exploration and cumulative exploratory efforts. The next sections discuss economic theory issues related to the optimal time path of exploratory effort (equation 2.5).

---

[10] A complete derivation and explanation of the variables of equation 2.5 is given in Appendix 2.1.

2.3.2 Exploration with uncertainty and learning

So far we have not yet explicitly addressed uncertainty in the process of exploration. Exploration of oil and gas is full of uncertainties. There is uncertainty over the future demand for the resource and uncertainty over the reserve base and its rate of recovery (Pindyck 1980). For example, uncertainty over the resource base includes the timing and magnitude of further discoveries of new reserves (Heal 1979). There is also uncertainty related to the size of reserves of a resource whose location is currently known. In terms of demand, there is the possibility that a close substitute of the resource would be developed in the future where the timing, discovery, and the scale of the substitute are uncertain. Stochastic models of resource exploration have been developed by Arrow and Chang (1978) and Deshmukh and Pliska (1980), in which discrete increments of reserve discoveries occur stochastically as a Poisson process in proportion to the level of exploration.

Firms reduce uncertainty related to the discovery process by continually updating their drilling decisions based on information gained in the previous year and/or neighboring drilling activities. This is modeled as exploration with learning (Quyen 1991). For example, suppose that a firm has drilled three wells on a prospect with the results that the first well is dry hole, the second shows a thin oil column, and the third is dry. Should the firm drill more wells to prove out the prospect and, if so, how many? Does the firm have sufficient information from the three wells already drilled to decide to abandon the opportunity? If the firm were to drill more wells, what fraction could be expected to be dry holes and what fraction could be expected to be oil bearing? These questions are answered using Bayesian updating procedures[11].

Bayesian learning is incorporated in equation 2.5. The optimal path of exploratory effort (w_{it}) is determined, among other factors, by additions to reserves, $d_{it} = f(w_{it}, A_{it-1})$. Since the exact amount and location of remaining reserves is not known, the number of successful wells discovered can be used to update information for the next round of exploration. For example, let the initial probability of finding an economic reservoir be p(ER), so that the probability of not finding an economic reservoir is p(NER) = 1 – p(ER) [12]. Four conditional probabilities can be considered:

1. The probability of drilling a dry hole given that an economic reservoir does exist, p(Dry|ER).
2. The probability of drilling a wet hole (oil or gas) given that an economic reservoir does exist, p(Wet|ER).
3. The probability of drilling a dry hole given that no economic reservoir exists, p(Dry|NER).

---

[11] Bayesian updating as applied to petroleum exploration is discussed in Lerche and Mackay (1999).

[12] In this case we are assuming that prior to any wells being drilled; firms must have made some exploration risk assessment to determine the probabilities using seismic data or other information.

4. The probability of drilling a wet hole given that no economic reservoir exists, p(Wet|NER). Note: there can still be some oil or gas even if it is too small to be economically exploited.

Firms update their probability that an economic reservoir exists given that *k* out of *n* wells resulted in finding a wet well as follows:

$$p(ER|kWet) = \frac{p(kWet|ER) \times p_{k-1}(ER)}{p(kWet|ER) \times p_{k-1}(ER) + p(kWet|NER) \times p_{k-1}(NER)} \qquad 2.6$$

The implication of the learning process for the empirical model is that firms drill wells in a certain region based on the information gained from previous drilling. It is expected that when the probability of success in an economic reservoir is high, more wells are expected to be drilled.

Quyen (1991) has developed a model of exploration with learning. His model incorporates learning based on spatial aspects of an exploration process using past geological information, specifically, past discoveries obtained from the region. In empirical studies, exploration with learning is often discussed together with the depletion effect of exploration. The depletion effect is defined as a decrease in additions to proven reserves as the number of explored regions increases (Quyen 1991). Cumulative drilling or cumulative discoveries are used as a proxy for the state of knowledge about the geology of the region (Uhler 1976, Siegel 1985, Iledare and Pulsipher 1999, Kemp and Kasim 2006).

The learning and depletion effects have opposing effects on exploration effort (Siegel 1985, Uhler 1976). Iledare and Pulsipher (1999) used cumulative drilling as an explanatory variable in their drilling equations for North and South Louisiana and found mixed results. They concluded that the negative coefficient of cumulative drilling in North Louisiana provides evidence of a resource depletion effect overpowering the learning effect, while the positive coefficient in the South could be due to the greater influence of the learning effect. Kemp and Kasim (2006) have also estimated an exploration efficiency (reserves discovered per well) equation using cumulative drilling as a proxy for technology and cumulative discoveries as a proxy for maturity or depletion effect. Based on their results, they point out that the positive effects of technology have amply compensated for the drag in exploration efficiency by maturity.

Based on the Bayesian updating procedures discussed above, learning effects are incorporated in our model using *lagged success rate* of wells as a proxy variable for learning. The study postulates that when a firm operating in a particular region drills an exploratory well, it collects information from all other firms in the neighborhood. This knowledge will allow the firm to choose drilling sites more wisely. The more successful wells are drilled at a specific region in the previous period, more drilling activities will occur in that region in the current period. Cumulative drilling and squared cumulative drilling are used to capture variables related to depletion effects.

2.3.3 Spatial exploration and regional models

So far we have not discussed how firms decide where to drill or how spatial entities are related to each other. In the previous section the process of learning was used to reduce uncertainties associated with exploration and reserve discoveries. Learning can also be used to explain interactions among neighboring spatial units in the process of exploration. Let us consider two types of exploratory activities. One is exploration at the intensive margin. In our case, the intensive margin can be defined as exploration effort within a given township or grid cell. The extensive margin is exploration effort in neighboring townships. Based on this classification, learning can take two forms. One is experience gained from additions to reserves in township i at time t. This can be referred to as learning in the intensive margin. The second type of learning is information accumulated from exploration activities in the neighboring townships. This can be referred to as learning at the extensive margin. Firms accumulate information on potential reserves, success rates, and discoveries in adjacent townships based on the notion that deposits tend to cluster. The concept of clustering of deposits explains that the discovery of a deposit immediately improves chances of there being another nearby deposit (Cairns 1990). As the exploration process unfolds, the information gathering process or learning is repeatedly revised each time exploration is carried out. In this case firms make their location choice based on comparison of expected returns at the intensive and extensive margins. This revision process continues until the entire region has been explored or when the exploration program is terminated.

Region-specific exploration: Deposits of natural resources are scattered around a vast region with different geologic and geographic characteristics. Most economic analyses of exploration and extraction are focused on developing an aggregate regional models overtime (e.g. Pindyck 1978, Peseran 1990, Kunce et al 2004). Kunce et al (2004) have developed an oil and gas exploration model for the United States to address the impact of environmental and land use regulations on drilling activities in the Wyoming Checkerboard. The assumption behind this model is that all resource sites or deposits in the U.S. exhibit similar geological characteristics. A major conclusion of their study is that drilling and future production of oil and gas is sensitive to changes in costs associated with environmental and land use regulations. Even though the authors have found policy relevant issues pertaining to drilling in Wyoming, they suggest that a more spatially disaggregated approach, at a sub-state or field level, would be superior to using state level data because of the considerable variability within states in drilling depths, sediment structure, and other cost determining factors. Similar issues were identified by Farzin (2001) who studied the impact of oil price on additions to U.S. proven reserves and by Lynch (2002), who critically reviewed several oil supply models.

The advantages of building regional models are many. Since region specific resources are deposited by similar geological processes, building regional models provides better empirical results to inform policy making processes. It is natural that policy measures concerning land impacts be directed to the region or even individual deposits. Since one of the objectives of this study is to understand energy sector drilling activities in wildlife habitats, developing a model of exploration tailored to a specific region would help to better identify conservation risks. The other advantage of regional models is that exploration functions are more likely to be well-specified for a defined region or play (Siegel 1985).

Recently, regional models of exploration have been getting more attention. Kemp and Kasim (2006) and Managi et al (2005) have developed regional offshore oil and gas exploration models for the UK Continental Shelf and the US Gulf of Mexico respectively. Iledare (1995) and Iledare and Pulsipher (1999, 2001) have done a number of region-specific studies on exploration and reserve additions in different parts of the United States. There have been several studies on exploration of oil and gas in Alberta. Uhler (1976, 1979) builds both theoretical and empirical models of petroleum exploration in Alberta. His results show that the largest reservoirs are discovered first and more discoveries are made as experience is gained in a particular play and there is an eventual reduction in exploration success as the basin becomes mature. Livernois (1988), and Livernois and Uhler (1987) have done similar studies using data from Alberta[13].

2.4 Empirical Model

An empirical model is developed by identifying the factors that affect drilling effort based on economic theory and the literature surveyed. Since functional forms for reserve additions, cost of exploration, and cost of extraction are not specified, equation 2.5 cannot be readily estimated. A common practice in the literature is to specify functional forms for reserve additions and cost functions and then insert the marginal product of exploration and marginal costs directly to equation 2.5 (Pindyck 1978, Pesaran 1990, and Kunce et al 2004). In this study we use the implicit function theorem to specify an equation of drilling effort as a generalized function of the factors that affect drilling. This function can be stated as:

---

[13] Studies by Livernois (1988) and Livernois and Uhler (1988) are mostly focused on estimating cost and supply function of oil and gas reserve additions and do not address environmental impacts.

$$F(w_{it}, d_{it}, A_{it-1}, R_{it}, P_t, \delta_t, \Lambda_t) \equiv 0 \qquad 2.7$$

where all the variables are as defined before. Assume that equation 2.7 is a continuous function and differentiable at w_{it}. According to the implicit function theorem, there is a unique value for drilling effort (w_{it}) such that equation 2.7 holds true. Drilling effort (w_{it}) can then be written as a function of the factors that affect drilling. These factors include: reserve additions, lag of cumulative wells drilled, reserves potential, price and technology. Hence equation 2.7 can be written as:

$$w_{it} = F\left(d_{it}, A_{it-1}, R_{it}, P_t, \delta_t, \Lambda_t / \beta_k, \varepsilon_{it}\right) \qquad 2.8$$

$\beta_1, \ldots \beta_k$ are unknown parameters to be estimated. ε_{it} are independently and identically distributed (iid) error terms for all *i* and t, with zero mean and constant variance (σ^2). Equation 2.8 is estimated using panel data models. Firms collect a substantial amount of information such as reserves in-place, success rates in nearby prospects, and seismic surveys before they start drilling exploratory wells in the next period. These data are not readily available. Instead, proxy variables that capture this information are used to estimate the model. For example, lagged cumulative number of wells and its square are good proxies to capture infrastructure variables and reserve depletion effects in a specific township in a given period. In particular, squared cumulative drilling is included to capture the depletion effect which will result in a downward trend of drilling activity when intensive drilling has already occurred in a certain township. According to the learning model developed in section 2.3, *lagged success rate* is used as a proxy variable to represent learning or updates on expected reserve additions. A trend variable is included to capture technical change and other variables that vary with time. For example, historical events related to the introduction of new regulations that affect exploration are captured in the trend variable. *Capacity utilization* rate captures technological constraints of drilling activities. Equation 2.8 can be written as an empirical model as follows:

$$w_{it} = F\left(S_{it-1}, A_{it-1}, A^{2}{}_{it-1}, P_t, T_t, U_t, D_t / \beta_k, \varepsilon_{it}\right) \qquad 2.9$$

where $S_{i(t-1)}$ refers to *lagged success rate* at township i in period t. $A_{i(t-1)}$ and $A^{2}{}_{i(t-1)}$ are the cumulative number of wells and its square respectively. P_t, T_t, CU_t ,and D_t represent price, technology, capacity utilization, and a dummy variable to capture the introduction of 3D seismic survey respectively. A detailed description of the variables is provided in the next chapter.

2.4.1 Spatial econometric issues

The inclusion of spatial effects in applied econometric models is typically motivated either on theoretical grounds, following from the formal specification of spatial interaction in an economic model, or on practical grounds due to peculiarities of the data used in an empirical analysis (Anselin 2002). On a theoretical ground, the fact that distance affects economic behavior is the main reason for an observation associated with a specific location being dependent on observations at neighborhood locations. In a regression framework, spatial autocorrelation occurs when the dependent variable or the error term of a regression function is correlated at each location with observations of the dependent variable or error terms at other locations. In other words, the magnitude of a decision variable for an economic agent depends on the magnitude of the decision variable set by other neighboring agents (Anselin 2002). For the case of oil and gas drilling in Alberta, this implies the number of wells drilled in a certain location will be correlated with the number of wells drilled at the neighboring locations. This is because firms collect information from neighbors before they further engage in drilling activities in their field of location. This provides the theoretical basis for taking spatial autocorrelation into account.

On the other hand, the motivation for applying a spatial econometric model may not be driven by formal theoretical concerns, but as the result of data problems. The scale and location of a data collection process may be arranged based on geographical location, which results in a systematic spatial pattern. For example, explanatory variables may be constructed by spatial interpolation to make their scale compatible with the dependent variable. Another often-encountered situation is when data on important variables are missing and those variables show spatial structure. A common characteristic of these data problems is that the error term in a regression model will tend to be spatially correlated. Ignoring this structure when it actually exists results in misspecification and estimation bias (Anselin, 1988). Preliminary tests based on a sample data from NE Alberta shows that drilling activities are spatially auto-correlated. Formal tests of spatial autocorrelations and results are discussed in chapter 3.

There are two options to address the problem of spatial autocorrelation. The first option is to take repeated random samples from different locations and estimate the equation (Chomitz and Gray 1996). This method is known as bootstrapping. This would help to reduce the spatial autocorrelation. The second option is to directly incorporate spatial autocorrelation in the model. In the present study the second option is used. Spatial autocorrelation is incorporated in the model using the spatial lag or/and spatial error models. The spatial lag model refers to the case where the dependent variable in a given location is affected by the variables in neighboring locations. The spatial error model refers to the case where the error terms across different spatial units are correlated (Anselin 1988). Specifications of these models in the context of spatial panel data are outlined in the next sub-section.

Spatial dependence or effects of location may also manifest as spatial heterogeneity. Spatial heterogeneity refers to the case where parameters estimated for the entire region may not adequately capture the process at a given sub-region. The problem of spatial heterogeneity is addressed in this study in two ways. In Alberta, a petroleum and natural gas license is issued for an initial term of two years if it is located in the Plains region, four years in the Northern region and five years in the Foothills region (DOE, 2005)[14]. At a regional level the problem is addressed by estimating parameters specific to the three geographic regions in Alberta. These terms take into account different geological and climatic conditions, topography, and access restrictions in the three regions. The regional models are estimated and formal statistical tests are conducted to test if the parameters are stable over the entire region. Spatial heterogeneity is also addressed at the observation level by estimating a fixed effects model. Fixed effects models are explained in the next sub-section.

2.4.2 Spatial specification of the model

There are two ways of specifying heterogeneity in panel data models. These are fixed effects and random effects approaches[15]. In this study, fixed effects panel data models that incorporate spatial dependence are used[16]. To simplify the notation, let the dependent variable and the independent variables specified in equation 2.9 be written as Y_{it} and X_{it} respectively. The effect of omitted variables that are peculiar to each spatial unit is represented by μ_i, and λ_t is a time-specific effect. Equation 2.9 can be specified as a fixed effects spatial lag panel data model as follows:

$$Y_{it} = \rho W\, Y_{it} + \beta X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \qquad 2.10$$

---

[14] A regional map of Alberta for the three regions is given in the next chapter.

[15] Different versions of panel data models that incorporate spatial components are discussed by Elhorst (2003, 2005).

[16] The fixed effects model is chosen based on Hausman's test for fixed versus random effects model for panel data. Model results reported in chapter 4, section 4.1 show that the fixed effects model is favored against the random effects model.

where W is a spatial weights matrix which is assumed to be constant over time and ρ is the spatial autoregressive coefficient for the spatial lag variable. It is assumed that λ_t, μ_i and ε_{it} are independent of each other and ε_{it} are independently and identically distributed (iid) error terms for all *i* and t, with zero mean and constant variance (σ^2). The spatial weights matrix (W) is constructed in such a way that all the diagonal elements are zero, spatial units that are neighbors to each other are given a value of one, and non-neighbors are given a value of zero. The matrix is then row standardized by dividing the sum of each row by the number of observations N. This specification is commonly called contiguity[17]. The three types of contiguity weights matrix in the classic example of a regular square grid layout are: the *rook* case (only common boundaries), the *bishop* case (only common vertices), and the *queen* case (both boundaries and vertices) (Anselin 2002). In this study the *queen* case is used where each township has a maximum of eight neighbors. In some cases townships along the borders of the province may have less then eight neighborhood townships.

Equation 2.10 can be specified as spatial fixed effects, a time-period fixed effects, or a spatial and time-period fixed effects model. The spatial fixed effects model refers to the case where spatial units are likely to differ in their background variables. These variables are usually space-specific and time-invariant variables that affect the dependent variable, but are difficult to measure or hard to obtain (Elhorst 2005). Omission of these variables leads to bias in the resulting estimates. One remedy for the bias is to introduce a variable intercept μ_i representing the effect of the omitted variables that are peculiar to each spatial unit considered. For a spatial fixed effects model λ_t is set to zero in equation 2.10. A time-period fixed effects model refers to the case where time periods differ in their background variables, which are usually time-specific spatial invariant variables that affect the dependent variable, but are difficult to measure or hard to obtain (Elhorst 2005). Generally, time period effects are assumed to be fixed and are justified by events such as policy interventions, structural breaks, sudden shocks, etc. In equation 2.10, μ_i is set to zero for a time-period fixed effects model. The spatial and time-period fixed effects model is a combination of the two models where both $\mu_i, \lambda_t \neq 0$.

---

[17] There are different ways of specifying the weights matrix. The two major specifications are distance- based and contiguity.

Equation 2.9 above can be specified as a fixed effects spatial error panel data model as follows:

$$Y_{it} = \beta X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad \text{where} \quad \varepsilon_{it} = \delta W \varepsilon_{it} + \upsilon_{it} \qquad 2.11$$

where δ is the coefficient of the spatial error variable and υ_{it} are independently and identically distributed (iid) error terms for all i and t, with zero mean and constant variance (σ_υ^2). μ_i, λ_t are spatial and time-period fixed effects respectively. Equations 2.10 and 2.11 are estimated in Matlab using routines of spatial panel data models developed by Elhorst (2005)[18]. These routines are specifically designed to estimate dynamic panel data models that include a spatially lagged dependent variable and spatial error autocorrelation. These equations are estimated using spatial data from Alberta oil and gas drilling activities from 1980 – 2004. Descriptive statistics and exploratory data analyses and tests of spatial autocorrelation are given in chapter three, and model results are given in chapter 4. Results obtained from equation 2.10 are used to forecast future exploration patterns and to test the energy sector's anticipation of environmental regulations in chapters five and six respectively.

---

[18] Detailed descriptions of estimation procedures of the spatial lag and spatial error models are given in Appendix 2.2.

Chapter 3 Data and Descriptive Analysis

3.1 Data sources and description

The study area covers all townships in Alberta where exploration of oil and gas is carried out. In total, there are approximately 7,200 townships in Alberta[19]. The study area is composed of 5,664 townships after Wood Buffalo, Jasper, and Banff National Parks are excluded. These parks are excluded based on the assumption that no exploration takes place in the parks. The study covers the time period from 1980 – 2004. In total, the panel data set is composed of 141,600 observations. Table 3.1 summarizes data required to estimate the empirical models discussed in chapter 2 and brief description of the data.

Table 3.1 Data required to estimate the empirical model and description of the variables

Variable Name	Brief description	Unit of measurement
Drilling effort	Number of wells drilled in a given township per year. Wells expressed per area of township.	$(\text{wells/km}^2)_t$
Cumulative wells	Total wells drilled per area of township from the beginning up to the specified year of observation.	$(\sum_{T=1}^{T} \text{wells/km}^2)_t$
Success rate	Proportion of wells that produce oil or/and gas in a given township per year	%
Price of oil	Average wellhead price of oil in a given year	$(\text{CND \$/bb})_t$
Price of gas	Average wellhead price of gas in a given year	$(\text{CND \$/GJ})_t$
Price of oil and gas	An index price for oil and gas in a given year.	Index number
Trend variable	Technology	Time
Capacity utilization rate	Number of active rigs divided by total rigs in a given year	%
3D seismic survey	Introduction of 3D seismic survey in 1990s	Dummy variable

---

[19] A description of the Alberta township system is explained in Appendix 1.1.

Well data: Data on the number of wells drilled per township is collected from GeoSCOUT™, a data base company based in Calgary. GeoSCOUT™ is a software package that provides comprehensive data in a GIS format for oil and gas exploration and production activities in Western Canada. Originally raw data is extracted on an individual well basis called well-tickets. For each well there is a location component and date of first drilling specified as drilling spud date. Once drilling is started it could take from a few days to several months to be completed. Based on this information the data was aggregated on a township basis per year. For example, all the wells drilled in 1981 in township *i* are summed up to get the number of wells drilled in that township. There are two approaches for setting the unit of measurement for the drilling effort variable. One is to use count of wells per township. The second option is to use density of wells per township. Count of wells is an integer value that shows the number of wells drilled in a township in a given year. Density of wells is the number of wells drilled in a township divided by the area of the township. Since the areas of each township are not equal, the density measure is preferred to the counts measure in order to normalize the size variation of townships. The other advantage of using density of wells is that density is a continuous variable which is convenient for estimation purposes. Hence, for the purpose of this study drilling effort is measured in terms of well density. For prediction or forecasting purposes, density of wells is multiplied by the area of the township to recover the actual counts of wells.

Cumulative wells: Cumulative number of wells is calculated by adding up the total number of wells drilled in each township. As the data base starts from 1980, the cumulative number of wells for this year was obtained by adding up all wells drilled in a given township since the beginning of drilling in Alberta up to the year 1980. For example, cumulative wells for the year 1981 are obtained by adding up all wells drilled since the beginning up to 1981 and so on. In some townships drilling activities were started in the 1920's. The square of cumulative wells are also included to capture depletion effect of reserves. The expected sign for cumulative wells is positive and square of cumulative wells is negative. The positive sign indicates that more wells are drilled in a given area due to clustering effects and the negative sign indicates depletion of reserves.

Success rate: Lagged *success rate* is used in the model as a proxy for the reserve discovery and learning process. For each well ticket the data shows whether a well drilled in a given year is producing oil or gas, or whether it was drilled and abandoned. *Success rate* is expressed as the proportion of wells producing oil or gas out of total wells drilled in a given township per year. The coefficient for *success rate* is expected to be positive. This indicates that more wells are expected to be drilled in township with higher success rates.

Price of oil and gas: Wellhead prices for oil and gas in Alberta are collected from the Alberta Department of Energy. Price of oil is measured in Canadian dollars per barrel and the price of gas is measured in Canadian dollars per gigajoule (GJ). The price of crude oil is determined by international market forces and is directly correlated with the reference price of WTI (West Texas Intermediate) (EUB 2005). The EUB (Alberta

Energy and Utilities Board) uses the WTI crude price as its benchmark for world oil prices, as Alberta crude oil prices are based on WTI netbacks to Edmonton. Netbacks are calculated based on the WTI at Chicago minus transportation and other charges from Edmonton to Chicago. Netbacks are adjusted for the exchange rate as well as crude oil quality (EUB 2005).

While crude oil prices are determined globally, natural gas prices are set in the North American market with little global gas market influence. However, natural gas prices are correlated with crude oil prices. A price index price for oil and gas is constructed using 1980 as a base year as follows:

$$Oilpriceindex_t = \frac{Oilprice_t}{Oilprice_{1980}} \qquad 3.1$$

$$GaspriceIndex_t = \frac{Gasprice_t}{Gasprice_{1980}} \qquad 3.2$$

$$PriceIndex_t = \frac{Oilpriceindex_t + Gaspriceindex_t}{2} \qquad 3.3$$

The price series is lagged to capture the impact of price expectations on drilling activities. The price coefficient is expected to have a positive coefficient.

Technology: A *time trend* variable is included as a proxy for technical change. The natural logarithm of time is used to make the variable more stationary. Dummy variables based on important dates when new innovations were introduced are also constructed. For example, the introduction of low-impact seismic survey methods (3D) and horizontal drilling in 1990 is expected to reduce the effort required for a successful well discovery. The expected sign for technology is an empirical question. A positive coefficient would indicate that new technology would help firms to drill more wells. However, a negative coefficient could indicate that firm would drill successful wells using seismic survey information and hence drilling effort will decrease through time.

Capacity utilization: Capacity utilization rate is calculated as the number of active rigs divided by the total number of rigs available in a given year in the province. This variable is used as a proxy variable for capital constraint. Data on the total number of available rigs and active rigs in a given year is collected from the Canadian Association of Petroleum Producers, CAPP (2006). An alternative variable to the *capacity utilization rate* would be to use total rigs in the province. The advantage of using total number of rigs is that this variable is not a function of price whereas capacity utilization could be correlated with price. However, the number of active rigs in the province is more representative of a capital constraint than all the rigs including the idle one. *Capacity utilization* is expected to have a positive coefficient. This shows that as the number of available rigs in the province increase there is a capacity of drilling more wells.

Other data: Infrastructure variables such as roads and pipelines are potentially good explanatory variables for drilling effort because they are associated with development costs. However, there are issues of endogeneity associated with the construction of pipelines and roads. Hence these variables are dropped to avoid complexity and model misspecification. Instead, density of the cumulative number of wells is used as a proxy for infrastructure.

3.2 Map of the study area

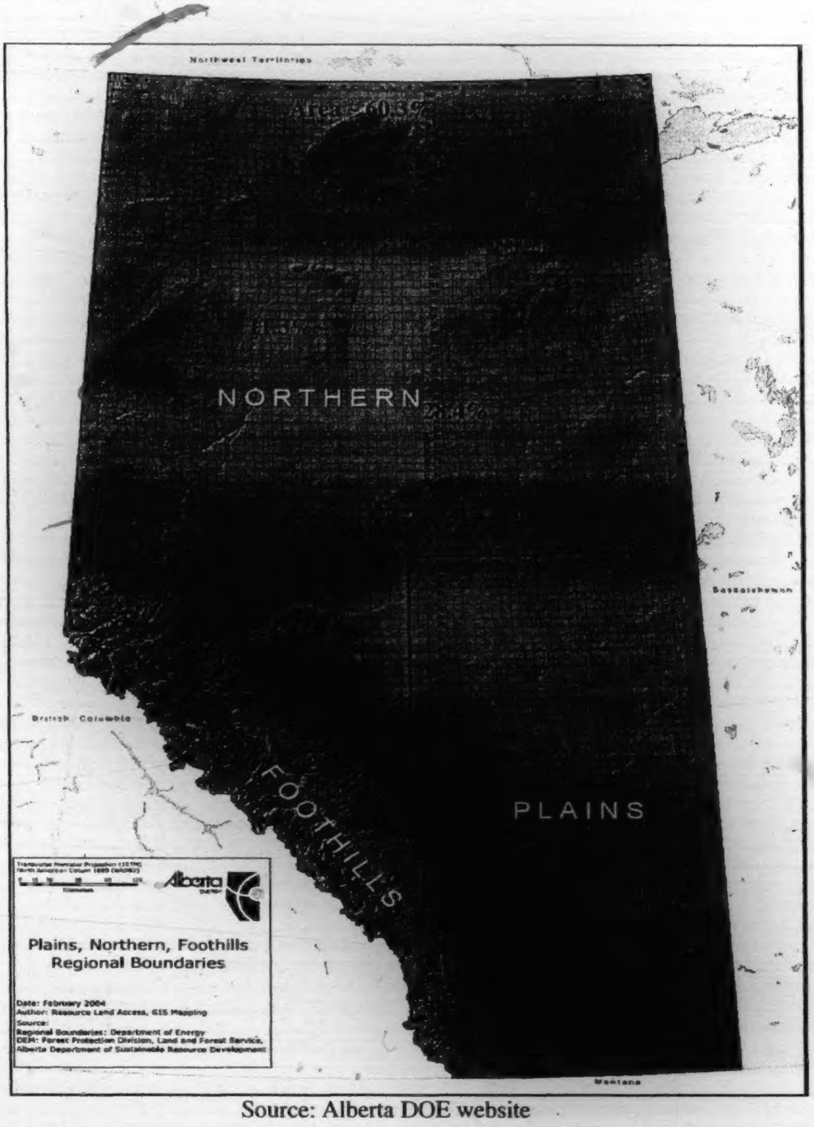
The study area covers the province of Alberta. This area is divided into three regions to capture the spatial heterogeneity issues discussed in the previous chapter. Figure 3.1 shows the three regions: the Northern, Plains, and the Foothills regions. These regions are heterogeneous in terms of costs of drilling and geographical and climatic conditions. For example, an initial oil and gas license is issued for two years if it is located in the Plains regions, four years in the Northern region and five years in the Foothills region. The minimum depth required to validate a license for the purpose of conversation from initial term to intermediate term[20] is 150 meters of depth in the Plains or Northern region and 300 meters in the Foothills region (DOE, 2005). The different policy regimes for the three regions imply cost differentials associated with drilling effort. Figure 3.1 shows that the Northern region covers 60.3 percent of the land and is composed of 3085 townships, excluding Wood Buffalo National Park. The Plains region covers 28.4 percent of the land and is composed of 2036 townships. The Foothills region covers only 11.3 percent of the land and is composed of 543 townships after the national parks along the Rocky Mountains are excluded. As discussed in the previous chapter, the sizes of all townships are not equal. Most townships are on average 93 km^2, although sizes vary. Some townships, due to the non-parallel nature of longitudinal lines over large north-south distances are as small as 0.6 km^2.

In terms of spatial scale, in the present study, the analysis is carried out at a township level. The other alternative would be to choose a smaller scale at a section or quarter of a section level. However, availability of data and the large number of sections of land with in the province of Alberta would limit the analysis.

---

[20] Intermediate term is defined as the second period of a license term, beginning with the day following the expiry of the initial term and continuing for five years, regardless of the location of the license (DOE, 2005).

Figure 3.1 Alberta regional boundary map


Northwest Territories
Area = 60.3%
NORTHERN
Saskatchewan
British Columbia
FOOTHILLS
PLAINS
Montana
Alberta
Plains, Northern, Foothills Regional Boundaries
Date: February 2004
Author: Resource Land Access, GIS Mapping
Source:
Regional Boundaries: Department of Energy
DEM: Forest Protection Division, Land and Forest Service, Alberta Department of Sustainable Resource Development

Source: Alberta DOE website

3.3 Descriptive statistics

Table 3.2 shows descriptive statistics of the variables discussed in the previous section. These statistics are based on data from each region in Alberta covering the time period of 1980 - 2004. Higher values of *well density* and cumulative *well density* in the Plains region show that there is a high concentration of oil and gas wells in this region than in the Northern or Foothills regions. Intensive drilling activities are carried out in this region due to huge gas reserves that lead to the drilling of more shallow wells On average the success rate of finding oil and gas wells is higher in the Plains region (26.9 percent) and it is lower in the Foothills region (4.4 percent). The average *success rate* in the Northern region is 12.6 percent. There are, however, large variances in the probability of finding an oil or gas well in the three regions. The rate of success varies from zero percent to almost 100 percent. The *capacity utilization* rate shows that on average 42 percent of the rigs are actively used. The minimum usage rate is 23 percent and the maximum is 63 percent. Some exploratory data analysis of wells is given in the next sub-section.

3.4 Exploratory data analysis

Figure 3.2 shows the total number of wells drilled in Alberta since 1980. These are aggregations of total wells drilled in Alberta for each year. A total of 273, 530 wells have been drilled in Alberta between 1980 and 2004. Tremendous increases in drilling activities have been observed in recent years, especially after the year 2000. This could be a reflection of an increase in demand for oil and gas both in the US and Canada, and an increase in the price of energy.

Table 3.2 Descriptive statistics of variables for the three regions in Alberta

Region	Variable	Mean	Stan.dev.	Minimum	Maximum
Northern	*Well Density*	0.012	0.046	0.00	2.206
	Cum. Well Density	0.164	0.337	0.00	9.398
	Cum. Well Density Squared	0.140	1.347	0.00	88.31
	Success Rate	0.126	0.295	0.00	0.99
Plains	*Well Density*	0.038	0.106	0.00	4.931
	Cum. Well Density	0.581	0.917	0.00	26.28
	Cum. Well Density Squared	1.180	9.504	0.00	691.07
	Success Rate	0.269	0.380	0.00	0.99
Foothills	*Well Density*	0.0025	0.010	0.00	0.220
	Cum. Well Density	0.035	0.088	0.00	1.225
	Cum. Well Density Squared	0.009	0.058	0.00	1.502
	Success Rate	0.044	0.189	0.00	0.99
All Regions	*Price Index of Oil and Gas*	1.455	0.559	0.878	2.670
All Regions	*Logarithm of Trend*	2.320	0.834	0.00	3.218
All Regions	*Capacity Utilization Rate*	0.42	0.02	0.23	0.63

Figure 3.2 Summary of wells drilled between 1980 and 2004.

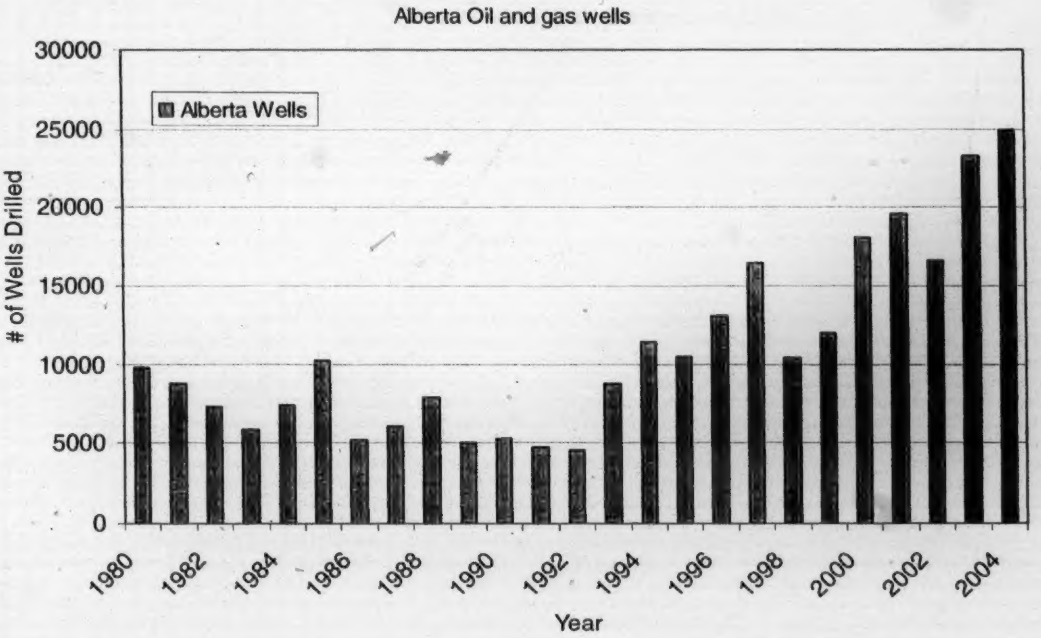
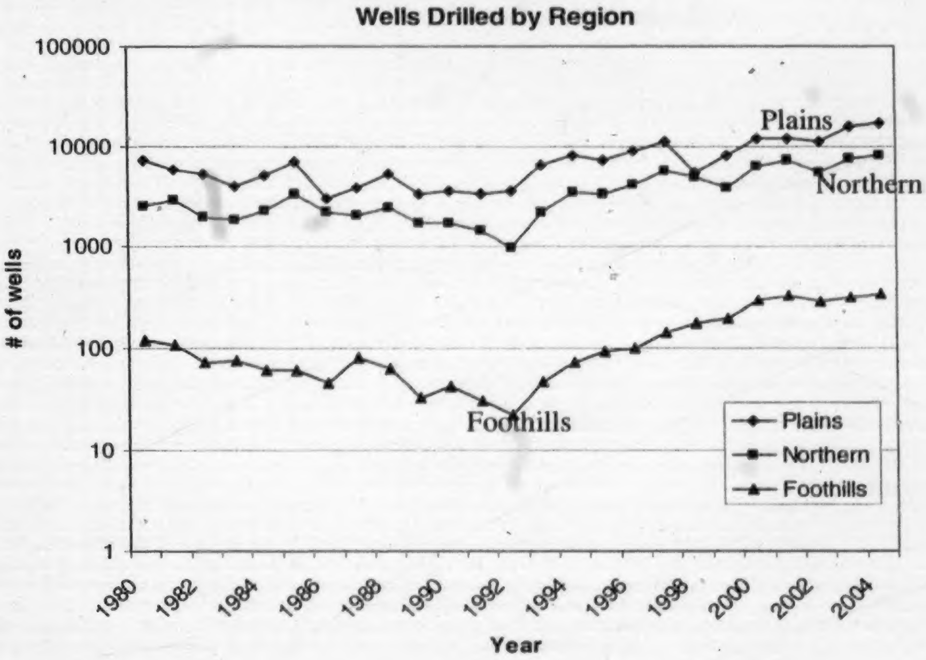


Figure 3.3 shows drilling activities classified by each region. The numbers are expressed in logarithms for ease of comparison in the graph. Given the size, geographical, and geological characteristic of the region, in absolute terms the numbers of wells drilled in the Foothills region are low compared with the other two regions. On average only 1% of the wells in Alberta are found in the Foothills region while 67% and 32% of the wells are drilled in the Plains and Northern regions respectively. Even though the Northern region covers 60% of Alberta's area, intensive drilling activities are observed in the Plains region. The difference in the rate of drilling could be attributed to the issues of spatial heterogeneity discussed in section 3.2.

Figure 3.3 Trend of regional drilling activities in Alberta



Figures 3.4 and 3.5 explain selected drilling activities in certain townships. For example, '406055' refers to area of land in Alberta located at the west of the fourth meridian, range 06 and township number 55. Figure 3.4 shows townships with intensive drilling activities where in some cases around 465 wells were drilled in a township per year. This particular observation is actually removed from the regression analysis because it is an outlier. Figure 3.5 shows a sample of townships with moderate drilling activities. In most cases frequent oscillations are observed, which makes modeling drilling activities on a township basis a difficult task. The next section discusses spatial autocorrelation tests.

Figure 3.4 Wells drilled through time for a sample of townships with intensive drilling

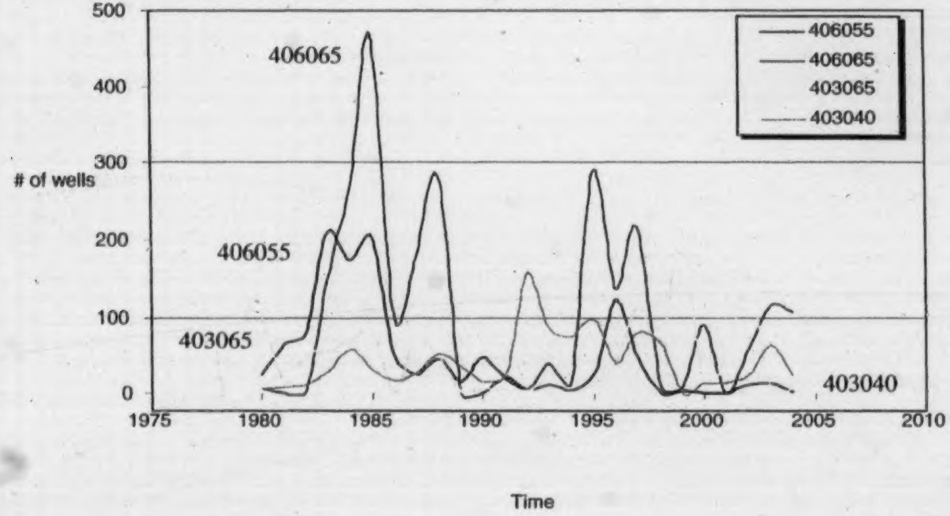
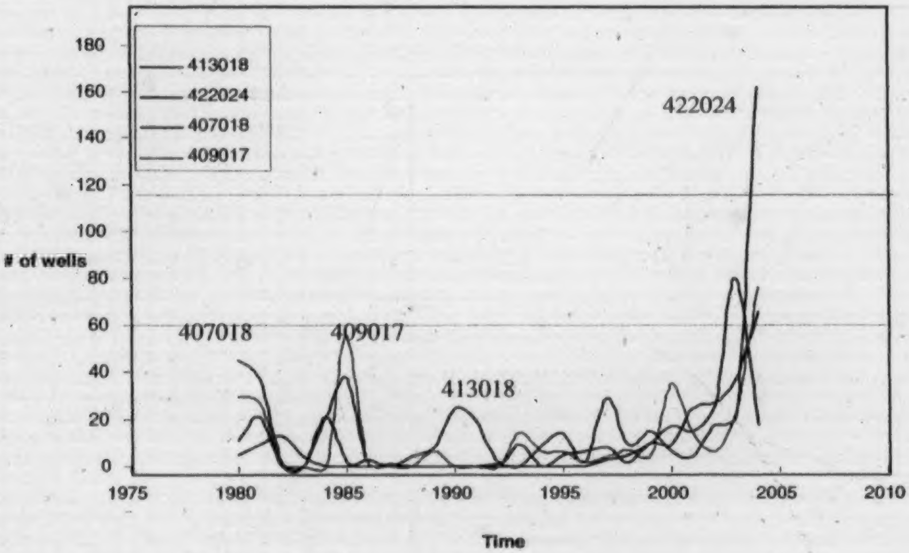


Figure 3.5 Wells drilled through time for a sample of townships with moderate drilling



3.5 Testing for spatial autocorrelation

The presence of spatial autocorrelation in the data is tested using Moran's I test. Spatial autocorrelation is the relationship between one observation and its neighbors across space. For example, the correlation between drilling effort in one location and its neighbors. The Moran's I test statistic for drilling effort can be written as:

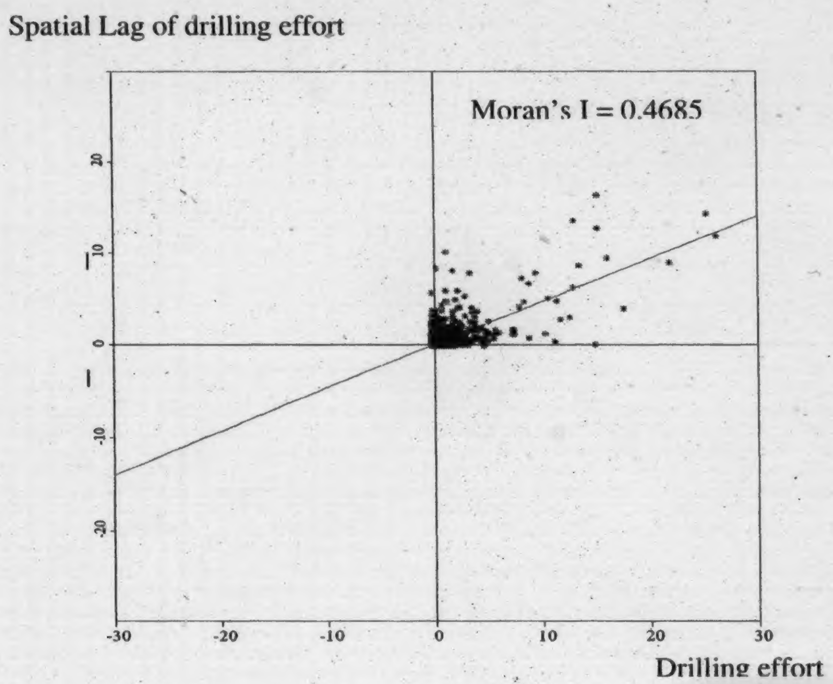
$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(Y_i - \overline{Y})(Y_j - \overline{Y})}{(\sum\sum W_{ij})\sum(Y_i - \overline{Y})^2} \qquad 3.4$$

where n is the number of observations and W is a weight matrix which takes a value of one for contiguous locations i and j and a value of zero otherwise. Y refers to drilling effort and $\overline{Y}$ is the mean value of Y. Values of Moran's I close to zero show no spatial pattern and close to +1 show strong positive spatial autocorrelation. For example, Figure 3.6 shows Moran's scatter plot of drilling effort in Alberta for the year 1982. This figure plots the density of wells against the spatially lagged density of wells[21]. The test statistic based on Moran's I value (0.4685 ($p < 0.001$) shows that there is positive spatial autocorrelation in the year 1982.

---

[21] Results for tests of spatial autocorrelation were obtained from data analyzed in GeoDA™ software.

Figure 3.6 Scatter plot of Moran's I for the year 1982



Similar tests were conducted for a sample of different years. Results for randomly selected years are presented in Table 3.3. These results show that there is evidence of positive spatial autocorrelation in the data.

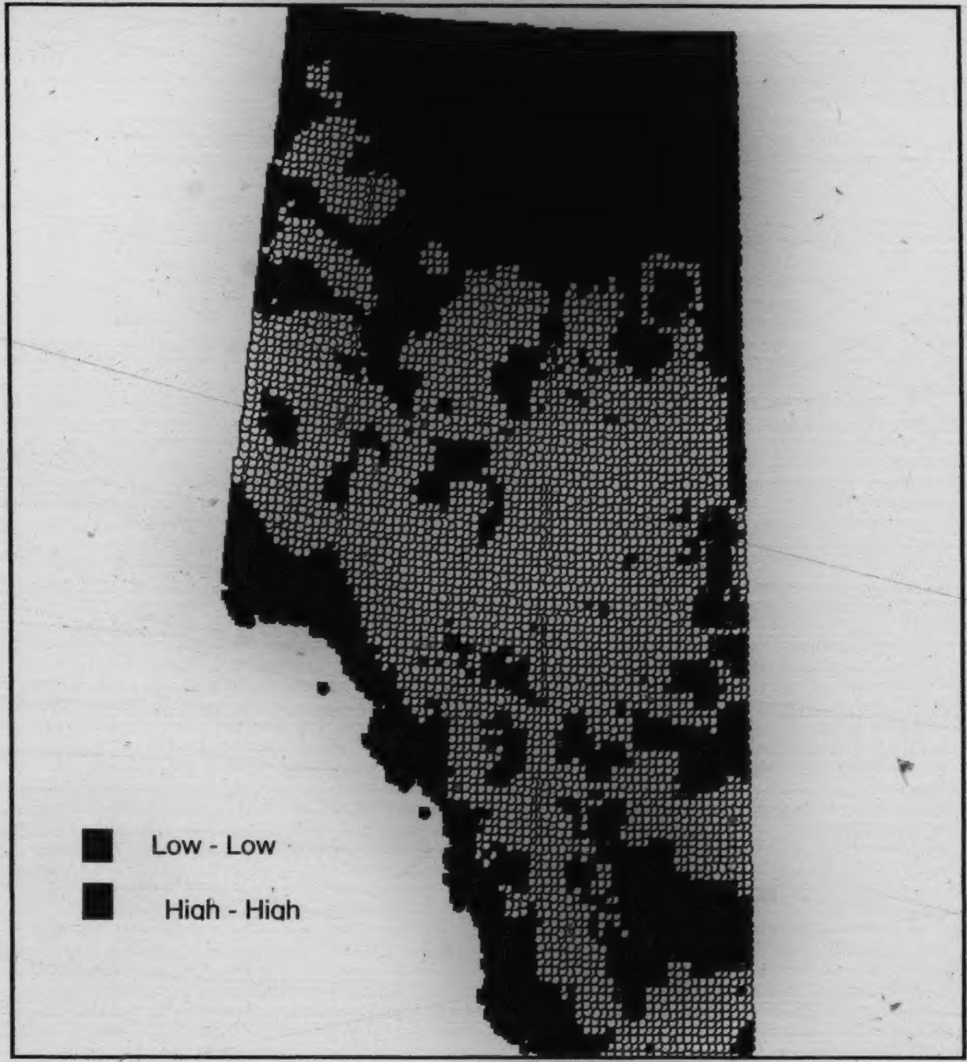
Table 3.3 Moran's I test for years 1985, 1992, 2000 and 2004.

Year	Moran's I value
1985	0.213 (0.001)
1992	0.327(0.001)
2000	0.407(0.001)
2004	0.436 (0.001)

Note: P-values are in parenthesis

Another approach of testing spatial autocorrelation is using Anselin's LISA (Local Indicator of Spatial Association), which can be seen as the local equivalent of Moran's I (Anselin 1995). For each location, LISA values show the tendency of high values of drilling densities to cluster with high values, creating hot spots of activity and the tendency of low values to cluster with low values creating cold spots. Spatial outliers are observed if high values of drilling densities are found closer to low values or low values closer to high values. Figure 3.7 shows a LISA map for spatial clusters and spatial outliers for the year 2000. The map shows that in 2000 hot spots are observed in the southeastern, eastern, and northeastern parts of Alberta. These regions include areas around Fort McMurray, Cold Lake, Athabasca, and Medicine Hat. The test results indicate that model estimation based on regional data is likely more robust than estimation based on aggregate data. The next chapter discusses results for different models using data from Alberta and the three regions with in Alberta.

Figure 3.7 Local indicator of spatial autocorrelation (LISA) map for the year 2000


■ Low - Low
■ High - High

Chapter 4 Model Results

This chapter summarizes the results for different versions of the model developed in chapter 2. The first section compares the results for spatial and non-spatial models. Section two deals with spatial lag and spatial error models. Analyses of regional spatial models within Alberta are discussed in section three and discussion and summary of the results are presented in the last sections.

4.1 Model estimation

Four models are estimated for Alberta. Table 4.1 shows the results for the spatial and aspatial models. These estimates are based on equations 2.14 and 2.15. The first two columns show estimates for random effects and fixed effects panel data models. The results indicate that the random effects and fixed effects models have similar coefficients in magnitude and sign. The coefficients for most of the variables have the expected sign. For example, the coefficients for cumulative wells, *success rate and first period lag of price* have positive and significant coefficients. The only difference between the random effects and fixed effects results are the coefficients for the second and third period price lags. The signs and the level of significance are different for these variables. For example, the third period price lag for the random effects model have a positive and significant coefficient while for the fixed effects model it has a negative and insignificant coefficient. Detailed interpretations of the results are given in the next sections.

Table 4.1 Results for aspatial and spatial models of drilling density in Alberta

	Aspatial model (Random effects)	Aspatial model (Fixed effects)	Fixed effects spatial lag model*	Fixed effects spatial error model*
Variable	Coef	Coef	Coef	Coef
Constant	-0.016 (-13.4)	-	-0.020 (-17.9)	-0.020 (-10.9)
Cum. Wells	0.047 (119.7)	0.045 (105.9)	0.050 (125.7)	0.040 (99.5)
Cum. Wells Squared	-0.001 (-17.9)	-0.001 (-15.9)	-0.001 (-16.4)	-0.001 (-16.4)
Lag. Success Rate	0.030 (57.8)	0.028 (53.8)	0.030 (59.9)	0.020 (45.4))
Lag Price-1 Period	0.004 (6.1)	0.005 (6.3)	0.004 (9.22)	0.005 (5.57)
Lag Price -2 Period	-0.000 (-0.18)	0.003 (0.22)	0.001 (1.16)	0.000 (0.16)
Lag Price-3 Period	0.002 (2.9)	-0.002 (-0.03)	0.002 (4.3)	0.002 (2.2)
Capacity Utilization	0.030 (14.9)	0.030 (21.6)	0.030 (20.0)	0.030 (12.37)
Time Trend	-0.001 (-2.4)	-0.030 (-1.02)	-0.001 (-2.58)	-0.001 (-1.59)
3D Seismic Dummy	-0.005 (-0.7)	-0.007 (-1.55)	-0.001 (-1.72)	-0.000 (-0.38)
Lag Dep. Variable (ρ)	-	-	0.060 (58.8)	-
Lag Spat. Error (δ)	-	-	-	0.490 (169.27)
N	141600	141600	141600	141600
R-Squared	0.22	0.28	0.29	0.34
Adj.R-Squared	0.22	0.25	0.29	0.34
Sigma^2	0.003	0.003	0.003	0.003
Log-Likelihood	230438.1	231158.5	233256.0	233473.8

Note: t-statistics in parentheses. * The fixed effects models are pooled model that do not take space and time into account. Models that take space and time into account are give in the next section.

The Hausman test (with a value of 259.6) for comparing fixed effects versus random effects panel data models show that fixed effects model is favored over random effects model. Hence, the rest of the spatial panel data models are estimated using fixed effects models. Intuitively, the use of fixed effects model helps to take into account unobservable variables in individual townships. Examples of these variables include differences in forest cover, underlying reserves, roads and related facilities.

The last two columns of Table 4.1 show spatial fixed effects panel data models. Most of the coefficients for the spatial lag and spatial error variables are significant and they have similar magnitudes. The coefficients for the spatial lag of the dependent variable (ρ= 0.06) and for the spatial lag of the error term (δ = 0.49) are both positive and significant at 5% level of significance. These results are consistent with the spatial autocorrelation tests discussed above. A comparison between the spatial and aspatial models shows that spatial models perform better. This is reflected in higher values of R-squared and log likelihood values. A comparison within the spatial models shows that the spatial error model has a higher R-squared and larger log likelihood ratio than the spatial lag model. However, the spatial lag model is supported by the theoretical underpinning of learning in geophysical exploration.

Assuming that the non-spatial fixed effects model is a restricted version of the spatial lag and/or the spatial error models, a formal Likelihood Ratio (LR) test is conducted to test if the coefficients of spatial dependence in the lag and error models are significantly different from zero. The results reported in Table 4.2 show that the coefficients are significantly different than zero.

Table 4.2 Likelihood Ratio (LR) test for spatial versus aspatial models

Models	LR = 2(ULR – RL)*	Critical Value
Spatial lag versus non-spatial	3396.0	10.83 (P = 0.001)
Spatial error versus non-spatial	6629.8	10.83 (P = 0.001)

* ULR and RL refer to log likelihood values for the unrestricted and restricted models respectively. Degree of freedom = 1.

4.2 Spatial lag versus spatial error models

In this section, different versions of the spatial lag and spatial error models are presented in Tables 4.3 and 4.4 respectively. In Table 4.3, column one shows a model estimated using spatial fixed effects, column two time period fixed effects, and column three both spatial and time period fixed effects. These specifications are based on the discussion in chapter 2 section 2.4. The results in Table 4.3 suggest that, the coefficients of the independent variables are similar. However, the model that takes into account both spatial and time period fixed effects performs better than the models that individually take into account only spatial or time period effects. This is based on higher values of R-squared and log likelihood values for the spatial and time period fixed effects model.

Table 4.4 gives the results for a spatial error model with spatial, temporal, and joint spatial –temporal fixed effects. The results are similar to those presented in Table 4.3. There are, however, a few differences among the spatial lag and spatial error models. For example, the coefficients for the *time trend* variable in the spatial error model are now positive for the spatial effects and for the spatial and time period effects models. In addition, the coefficients for 3D seismic variable are significant for the spatial effects and for the spatial and time period effects models.

Table 4.3 Results for the spatial lag models of drilling density in Alberta

	Spatial fixed effects	Time period fixed effects	Spatial and time period fixed effects
Variable	Coef	Coef	Coef
Constant	-	-	-
Cum. Wells	0.040 (49.5)	0.050 (125.8)	0.040 (50.1)
Cum. Wells Squared	-0.003 (-54.8)	-0.001 (-16.5)	-0.003 (-55.1)
Lag. Success Rate	0.030 (53.5)	0.030 (60.1)	0.030 (53.8)
Lag Price-1 Period	0.004 (10.58)	0.004 (12.5)	0.005 (15.35)
Lag Price -2 Period	0.000 (0.92)	-0.001 (-2.4)	0.000 (0.14)
Lag Price-3 Period	0.002 (4.87)	0.002 (6.7)	0.002 (7.6)
Capacity Utilization	0.030 (23.1)	0.030 (28.1)	0.030 (31.8)
Time Trend	-0.001 (-1.73)	-0.001 (-4.2)	-0.001 (-2.21)
3D Seismic Dummy	0.000 (0.67)	-0.001 (-2.59)	0.001 (0.42)
Lag Dep. Variable (ρ)	0.060 (57.9)	0.060 (57.7)	0.060 (58.5)
N	141600	141600	141600
R-Squared	0.32	0.24	0.32
Adj.R-Squared	0.29	0.24	0.29
Sigma^2	0.003	0.003	0.003
Log-Likelihood	237510.2	237762.8	237511.5

Note: t-statistics in parentheses

Table 4.4 Results for the spatial error models of drilling density in Alberta

	Spatial fixed effects	Time period fixed effects	Spatial and time period fixed effects
Variable	Coef	Coef	Coef
Constant	- -	- -	- -
Cum. Wells	0.020	0.04	0.020
	(19.3)	(100.1)	(20.3)
Cum. Wells Squared	-0.002	-0.001	-0.002
	(-43.2)	(-16.6)	(-43.8)
Lag. Success Rate	0.020	0.020	0.020
	(39.1)	(45.5)	(39.5)
Lag Price-1 Period	0.006	0.005	0.006
	(7.89)	(12.5)	(15.9)
Lag Price -2 Period	0.001	-0.001	0.001
	(1.71)	(-1.42)	(2.95)
Lag Price-3 Period	0.003	0.002	0.003
	(2.89)	(5.6)	(7.27)
Capacity Utilization	0.040	0.030	0.040
	(14.75)	(23.4)	(27.3)
Time Trend	0.001	-0.001	0.001
	(0.71)	(-2.1)	(2.1)
3D Seismic Dummy	0.001	-0.001	0.001
	(1.75)	(-1.46)	(2.56)
Lag Spat. Error (δ)	0.500	0.490	0.500
	(169.53)	(167.9)	(169.5)
N	141600	141600	141600
R-squared	0.42	0.34	0.41
Adj. R-squared	0.39	0.34	0.39
Sigma^2	0.003	0.003	0.003
Log-Likelihood	236511.0	237762.3	248511.5

Note: t-stat in parenthesis

Comparing spatial lag and spatial error models is not a straightforward issue in the spatial econometrics literature (Anselin 2002). Spatial econometric models can contain both a spatially lagged dependent variable and spatially auto-correlated error terms. However, models that combine both cases are rarely used in practice because of problems of identification. The common practice is either to choose a model that is theoretically meaningful or to choose the model with a higher R-squared value (Anselin 2002). In the present study, even though the spatial error model has a larger R-squared value, theoretically the spatial lag model is more appealing. This is because when firms are planning to drill oil and gas wells in a given township, they take into account drilling information from surrounding townships. The spatial lag model captures this phenomenon by including the average number of wells in the surrounding townships.

4.3 Regional models

Different models are estimated for the three regions to capture regional differences in terms of length of contract for petroleum and natural gas licenses, geographical location, and geological characteristics of the areas. Results for regional spatial lag models are reported in Table 4.5. For each region two different versions of the model are estimated: a pooled fixed effects model and a fixed effects model that takes into account both spatial and time period effects. A comparison between these models shows the model that takes space and time into account performs better than the pooled models. For example, for the Plains region the value of R-squared increased from 0.22 to 0.30 for the fixed effects model[22].

---

[22] For the Northern region the value of R-squared increased from 0.18 to 0.25 and for the Foothills region it increased from 0.25 to 0.30.

Table 4.5 Results for regional spatial lag models

	Northern		Plains		Foothills	
	Pooled model	Spatial /time effects	Pooled model	Spatial /time effects	Pooled model	Spatial /time effects
Variable	Coef	Coef	Coef	Coef	Coef	Coef
Constant	-0.010	-	-0.040	-	-0.002	-
	(-13.1)		(-13.9)		(-6.0)	
Cum. Wells	0.050	0.050	0.040	0.030	0.060	0.070
	(80.1)	(42.6)	(56.6)	(17.2)	(30.2)	(18.5)
Cum. Wells Squared	-0.003	-0.005	-0.001	-0.003	-0.040	-0.060
	(-16.4)	(-21.1)	(-6.1)	(-26.2)	(-14.7)	(-13.1)
Lag. Success Rate	0.020	0.020	0.040	0.040	0.010	0.010
	(38.7)	(37.2)	(34.2)	(30.8)	(29.8)	(24.7)
Lag Price-1 Period	0.002	0.003	0.008	0.010	0.001	0.001
	(6.7)	(9.5)	(5.7)	(11.5)	(4.7)	(8.2)
Lag Price -2 Period	0.001	0.001	-0.003	-0.000	0.001	0.000
	(1.7)	(4.1)	(-1.8)	(-0.2)	(2.5)	(1.8)
Lag Price-3 Period	-.001	-.000	0.007	0.009	-.000	-.000
	(-1.1)	(-0.9)	(5.0)	(8.4)	(-0.3)	(-0.6)
Capacity Utilization	0.020	0.020	0.020	0.080	0.003	0.003
	(13.9)	(21.5)	(15.6)	(26.0)	(4.2)	(5.1)
Time Trend	0.000	-0.000	-0.003	-0.005	-0.001	-0.001
	(0.4)	(-0.4)	(-3.1)	(-0.7)	(-2.9)	(-2.1)
3D Seismic Dummy	-0.001	-0.001	0.000	0.003	0.001	0.000
	(-3.3)	(-3.9)	(0.1)	(2.9)	(2.5)	(1.3)
Lag Dep. Variable (ρ)	0.040	0.040	0.060	0.060	0.030	0.010
	(9.0)	(9.1)	(31.8)	(32.1)	(7.6)	(7.2)
N	77125	77125	50900	50900	13575	13575
R-Squared	0.18	0.25	0.22	0.30	0.25	0.30
Adj.R-Squared	0.18	0.22	0.21	0.27	0.25	0.27
Sigma^2	0.001	0.001	0.009	0.009	0.0001	0.0001
Log-Likelihood	134950	138270.7	47731.2	50774.4	44593.9	45113.0

Note: t-stat in parenthesis

A Likelihood Ratio (LR) test is used to test for regional heterogeneity by comparing whether there are statistically significant differences among the coefficients of the explanatory variables for Alberta and the three regional models. The model for Alberta is a restricted version of the three separate models for each region. The hypothesis for the regional comparisons of the coefficients can be formally written as follows:

$$H_0 : \beta_A = \beta_N = \beta_P = \beta_F \qquad 4.2$$

$$H_1 : \beta_A \neq \beta_N \neq \beta_P \neq \beta_F$$

where βs' indicate vectors of the coefficients for the explanatory variables for Alberta, Northern, Plains, and Foothills models respectively.

Results for the LR tests are given in Table 4.6[23]. The results show that values of the LR are greater than the critical values. This implies regional heterogeneity. This is true for both the pooled and spatial and time period fixed effects models. Therefore the regional models are preferred for forecasting purposes.

Table 4.6 Results of Likelihood Ratio (LR) test for the spatial heterogeneity hypothesis: Spatial lag model

Models	LR = 2(ULR – RL)*	Null hypothesis
Pooled model	237.8	13.82 (P = 0.001)
Spatial/time effects model	313.2	13.82 (P = 0.001)

* ULR and RL refer to log likelihood values for the unrestricted model (sum of three regional models) and the restricted model (Alberta model) respectively. Degrees of freedom = 2.

---

[23] The likelihood value for the unrestricted model is the sum of log likelihood values for the three regions, and for the restricted model it is the log likelihood value for Alberta.

A closer look at the coefficients of the explanatory variables for each region reveals a number of interesting interpretations. For example in Table 4.5, the coefficient on cumulative number of wells is positive and significant and the coefficient on the square of cumulative number of wells is negative and significant. The positive coefficient shows that more wells are expected to be drilled in townships where there are more reserves. This is also an indication of clustering of deposits as discussed by Cairns (1990). The negative coefficient on the squared term is an indication of depletion effect of exploration through time. These results are consistent with the findings of other studies (Uhler, 1976, Siegel, 1985, Iledare and Pulsipher, 1999). The only difference is the way the cumulative number of wells is used in the equations. For example, Siegel (1985) used cumulative wells to capture both geological knowledge and depletion of drilling sites and Uhler (1976) used cumulative discoveries to capture both learning and depletion effects. In the present study, cumulative wells are used to capture the effects of reserves and other infrastructure variables and square of cumulative wells are used as a proxy for depletion effect. *Success rate* is used to capture learning effects. A positive and significant coefficient of lagged *success rate* is an indication of a learning effect in which more wells are drilled in areas where there are high rates of successful discoveries.

The first, second, and third period lags of wellhead price of oil and gas were included in the equations to capture expectation formation. Coefficients for the first and the third period lag are positive and significant while the coefficient for the second period is not significant. Positive price coefficients show that firms tend to drill more wells when they anticipate that the price of oil and gas is expected to increase. Discussion of price elasticities is given in the last section of this chapter. The *capacity utilization* rate was included to capture capacity constraints in drilling. This coefficient is positive and significant which implies that more wells are expected to be drilled as the capacity in the province increases. The time trend, a proxy variable for technical change, has a mixed sign. In some models the negative coefficient could be an indication of the introduction of new technologies that enable firms to locate new reserves of oil and gas easily and hence drill fewer wells. The 3D seismic variable is a dummy variable that captures the introduction of 3D seismic survey technology in 1990s. 3D seismic survey helps firms to easily locate reserves of oil and gas under the surface. The sign of the coefficient for the 3D survey variable was expected to be positive. However, the results show that it has a mixed sign in the three regions.

The results for the spatial and time period fixed effects model in Table 4.5 show that the coefficient for lagged *success rate* (0.01) is lower in the Foothills region than in the other two regions. This indicates that that it takes more effort to locate oil and gas wells in the Foothills regions than in the other regions. Furthermore, the coefficients for the temporal variables (lagged prices, capacity, and 3D seismic) are lower in the Foothills regions than in the other regions. This could indicate that firms would prefer to drill in the Northern or Plains regions than in the Foothills region. This is because limited infrastructure and the environmental sensitivity of the region could have made exploration activities more of a challenge than in the Plains or Northern regions. It may also be an indication of the depth of the deposits that makes deep drilling difficult.

A comparison between the Plains and the Northern regions show that the coefficients of the lagged *success rate* and the temporal variables are higher in the Plains than in the Northern region. For example, the coefficient of the first period lagged price is higher in the Plains (0.01) than in the Northern region (0.003). Similarly, the coefficient of lagged *success rate* is higher in the Plains (0.04) than in the Northern region (0.02). This shows that firms would prefer to drill in the Plains region compared with the Northern region.

Results for the estimates of the spatial error model for the three regions are given in Table 4.7. For each region two different versions of the model are estimated: a pooled fixed effects model and fixed effects model that takes into account spatial and time period effects. A comparison between these models shows that the model that takes into account the spatial and time period effects performs better than the pooled model. For example, for the Northern region the value of R-squared increased from 0.27 to 0.33 when the pooled model was re-estimated taking into account spatial and time period effects. For the Plains region the value of R-squared increased from 0.33 to 0.41 and for the Foothills region it increased from 0.27 to 0.32. These results are similar to the regional spatial lag model presented Table 4.5.

Table 4.7 Results for regional spatial error models

	Northern		Plains		Foothills	
	Pooled model	Spatial /time effects	Pooled model	Spatial /time effects	Pooled model	Spatial /time effects
Variable	Coef	Coef	Coef	Coef	Coef	Coef
Constant	-0.009 (-8.5)	-	-0.040 (-8.1)	-	0.003 (-5.6)	-
Cum. Wells	0.060 (78.6)	0.060 (38.5)	0.040 (46.2)	0.005 (13.3)	0.060 (30.2)	0.060 (18.1)
Cum. Wells Squared	-0.004 (-25.7)	-0.007 (-28.8)	-0.001 (-5.8)	-0.002 (-20.8	-0.040 (-14.8)	-0.050 (-12.8)
Lag. Success Rate	0.010 (30.1)	0.010 (29.1)	0.030 (25.6)	0.030 (22.8)	0.010 (29.2)	0.010 (24.3)
Lag Price-1 Period	0.002 (4.6)	0.002 (9.2)	0.010 (4.0)	0.010 (12.7)	0.001 (4.4)	0.002 (8.0)
Lag Price -2 Period	0.001 (1.2)	0.001 (4.6)	-0.001 (-0.6)	0.003 (2.7)	0.000 (2.2)	0.001 (1.7)
Lag Price-3 Period	-.001 (-1.1)	-.000 (-1.0)	0.007 (2.7)	0.009 (7.6)	-0.000 (-0.3)	-.000 (-0.5)
Capacity Utilization	0.020 (9.1)	0.020 (19.4)	0.080 (9.4)	0.090 (23.2)	0.003 (3.9)	0.003 (4.7)
Time Trend	0.000 (0.2)	0.000 (0.1)	-0.003 (-1.7)	0.002 (2.7)	-0.001 (-2.6)	-0.001 (-1.9)
3D Seismic Dummy	-0.001 (-2.2)	-0.001 (-2.9)	0.001 (0.4)	0.005 (3.8)	0.001 (2.3)	0.001 (1.2)
Lag Spat. Error (δ)	0.420 (99.8)	0.410 (99.6)	0.500 (93.4)	0.520 (95.5)	0.260 (6.2)	0.240 (7.1)
N	77125	77125	50900	50900	13575	13575
R-squared	0.27	0.33	0.33	0.41	0.27	0.32
Adj. R-squared	0.27	0.30	0.33	0.38	0.27	0.29
Log-Likelihood	137410.9	140666.7	50481.6	53646.4	44725.6	45220.5

Note: t-statistics in parentheses.

Similar to the results given in Table 4.7, a Likelihood Ratio (LR) test is also conducted for the spatial error model to test if there are statistically significant differences among the coefficients of the explanatory variables for Alberta and the three regional models. The procedures used to formulate the null and alternative hypotheses are similar to the spatial lag model. Results of the tests are given in Table 4.8. The results show that the null hypothesis of homogenous coefficients across the regions is rejected for both models.

Table 4.8 Results of Likelihood Ratio (LR) test for the spatial heterogeneity hypothesis: Spatial error model

Models	LR = 2(ULR – RL)*	Null hypothesis
Pooled model	345.8	13.82 (P = 0.001)
Spatial/time fixed effects model	811.9	13.82 (P = 0.001)

* ULR and RL refer to log likelihood values for the unrestricted model (sum of three regional models) and the restricted model (Alberta model) respectively. Degrees of freedom = 2.

A comparison between regional spatial lag and spatial error models show that spatial error models have higher R-squared and log likelihood values than the spatial lag models for each region. For example, in the Plains region a model estimated using a spatial error specification has a higher R-squared value (0.41) than a model estimated using a spatial lag model (0.30). This could indicate that spatial error models perform better than the spatial lag models. However, as discussed in section 4.3, the spatial lag model is more theoretically sound to adopt for each region. Trendle (2006) and Patton and McErlean (2003) are two examples in the literature who adopted the spatial lag model based on theoretical grounds rather than the spatial error model which performed statistically better in terms of higher R-squared and log likelihood values.

4.4 Discussion of results

The two most commonly cited spatial models, the spatial lag and spatial error models, are used to estimate the oil and gas exploration model. Three versions of the models: spatial fixed effects models, time period fixed effects models, and spatial and time period fixed effects models are estimated for Alberta and the three regions in Alberta[24]. Among these specifications, models that take into account both spatial and time period effects perform better than the others. A comparison between the spatial lag and spatial error models show that statistically the spatial error model performs better than the spatial lag model. However, based on theoretical grounds the spatial lag model is adopted in the next chapters for forecasting and analysis of environmental regulation purposes. In addition to the theoretical motivation discussed in chapter 2, a number of reasons could also be given why the spatial lag model is more appealing. One reason is that oil and gas firms would prefer to explore in areas where they have access to infrastructure such as pipelines and other facilities. The other reason is that reserves of oil and gas tend to cluster in a certain area which leads to intensive drilling activities. Given these facts, the spatial lag model is more applicable because it captures these activities by including the average drilling activities on neighboring townships on the right hand side of the equation (spatial lag of the dependent variable).

The results obtained from the estimation of the spatial panel data models show that a number of spatial and temporal explanatory variables are included in the model to explain the behavior of drilling effort. Most of the coefficients were found to be statistically significant and their signs are also as expected. The price of oil and gas is one of the temporal variables included in the model. First, second and third period lags of prices are included to capture price expectations. The results show that the first period lag of price is always positive and significant in all the models. However, the second period is mostly not significant and the results are mixed for the third period price lag. These results suggest that firms focus on the first period price lag. In other words, long term price expectations are not observed in the data.

---

[24] Definition and specifications of these models is given in chapter 2, section 2.4.2.

A more meaningful interpretation of the price coefficient is to examine the price elasticity of drilling effort. A formula for calculating elasticity is given by: $\varepsilon = {}^{\partial Y}\!/_{\partial P} \times {}^{\overline{P}}\!/_{\overline{Y}}$ where ε is the price elasticity, Y and P are drilling effort and price respectively, and $\overline{Y}$ and $\overline{P}$ are average values of drilling and price. The marginal effect of price (${}^{\partial Y}\!/_{\partial P}$) from the spatial error model is β where as the marginal effect of price from the spatial lag model is calculated as $\beta/(1-\rho)$, where β is the coefficient of the first period price lag and ρ is the coefficient of the spatial lag of drilling effort[25]. A summary of price elasticities for the three regions in Alberta is given in Table 4.9. The results show that the price elasticities are all inelastic. This implies that in the short run drilling is not very responsive to changes in the price of oil and gas. A comparison between the regions shows that drilling effort is relatively more inelastic in the Northern and Plains region than in the Foothills region. This could indicate that the percentage increase of drilling activities to a change in price is relatively more responsive in the Foothills regions because this region has relatively more unexplored land.

---

[25] The derivation of marginal effects and elasticities in the context of a spatial lag and spatial error models is explained in Kim et al (2003).

Table 4.9 Summary of price elasticities of drilling effort for the spatial lag and spatial error models

	Spatial lag model			Spatial error model		
	$\partial Y/\partial P = \beta/(1-\rho)$	$\bar{P}/\bar{Y}$	ε	$\partial Y/\partial P = \beta$	$\bar{P}/\bar{Y}$	ε
Northern region	0.0031	121.25	0.38	0.002	121.25	0.24
Plains region	0.0106	38.29	0.41	0.01	38.29	0.38
Foothills region	0.0010	582.0	0.59	0.002	582.0	1.16

Further analyses of temporal price elasticities of drilling effort are also performed. A graphical representation for Alberta and the three regions are given in Figures 4.1 to 4.4. The graphs show that price elasticities were higher prior to 1992 for all regions. Recent trends show that price elasticities are closer to the average values. The range is between 0.3 and 0.5 for all regions except for the Foothills region. Recent temporal elasticities for the Foothills regions vary between 0.5 and 0.78. The decrease in price elasticities through time could be a reflection of the maturity of the field and limited unexplored area remaining. According to the survey study by Dahl and Duggan (1998), price elasticities of drilling effort vary among different studies. For example, they reported price elasticity of 0.90 according to the study by Kolb and 0.48 according to a study by Al Shami. Even though the price elasticities are closer in magnitude to the present study, these results may not be directly comparable to the present study. This is because the specification of the model and the variables included in the equations are different from the previous studies.

Figure 4.1 Temporal price elasticity of drilling: Alberta

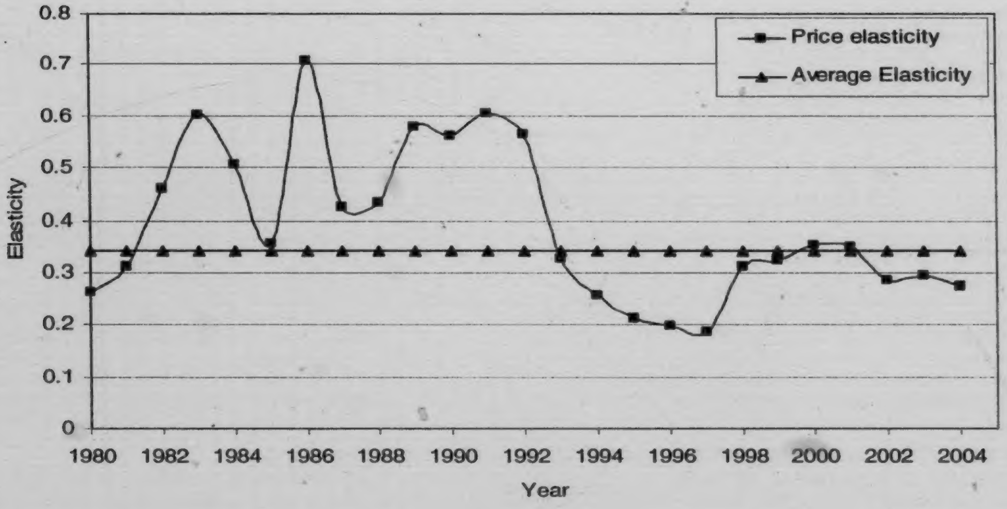


Figure 4.2 Temporal price elasticity of drilling: Northern region

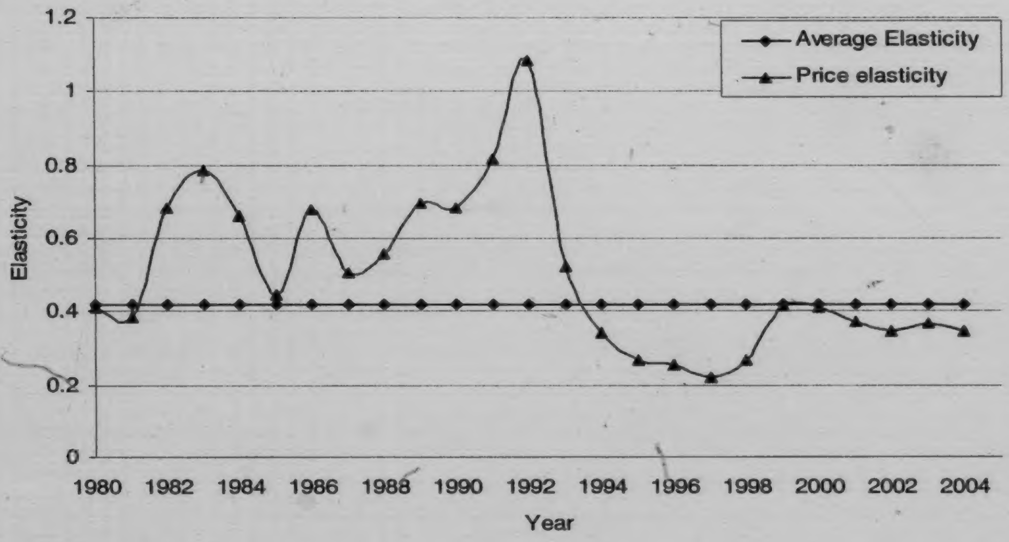


Figure 4.3 Temporal price elasticity of drilling: Plains region

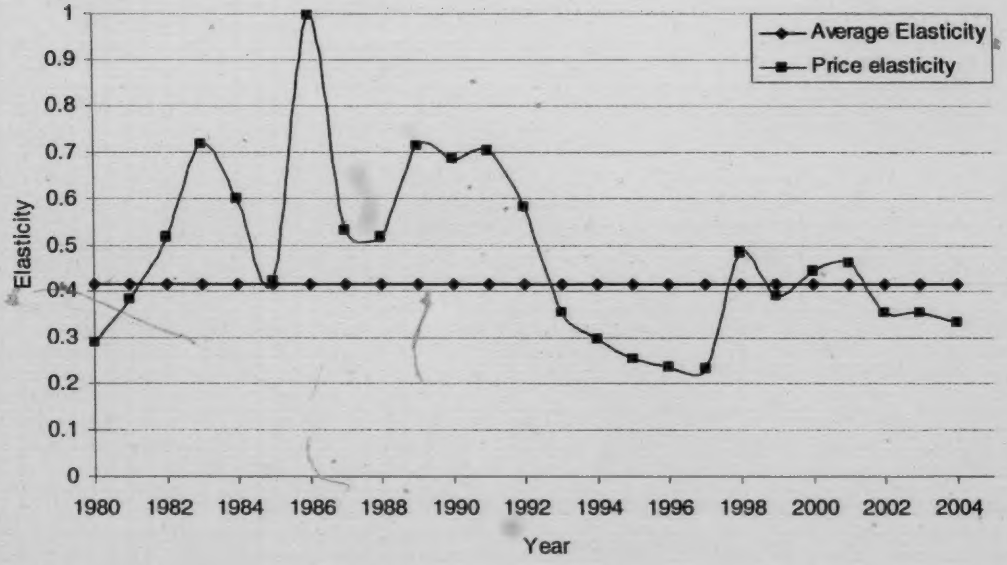
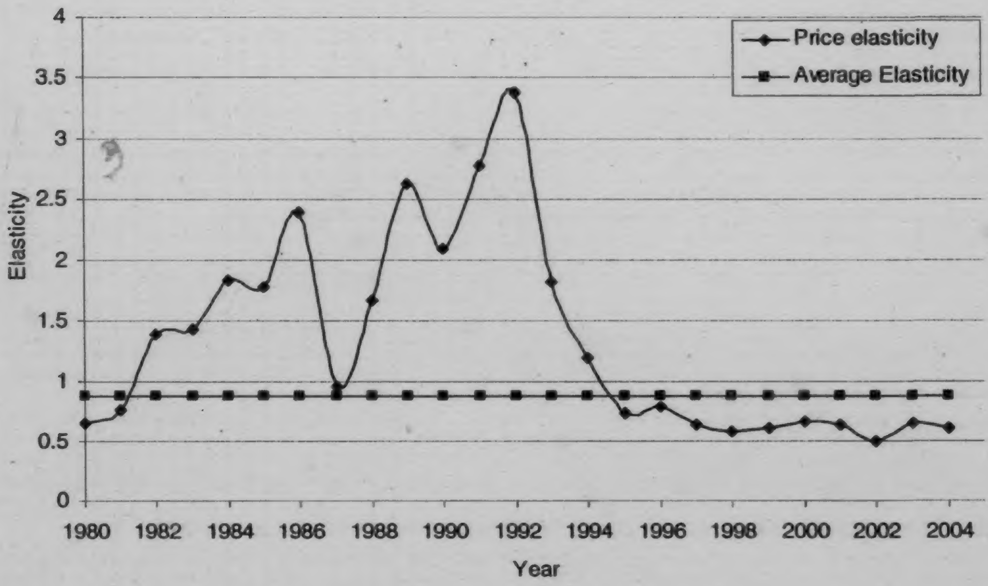


Figure 4.4 Temporal price elasticity of drilling: Foothills region



There are a number of major stylized facts or characteristics of non-renewable resources discussed in the economics literature. One example is the issue of uncertainty in the exploration of resources in terms of the location, size and quality of the resource (Attanasi, 1979). Other examples are depletion effects and clustering of deposits (Cairns 1990). The empirical models estimated in this paper relate to these stylized facts in the following way. We have discussed in chapter 2 that firms reduce uncertainty by collecting information on drilling activities from the previous years. To this effect, firms would deal with uncertainties by restricting their exploration activities to a small set of areas where they have previous experience or plans for prolonged exploration programs. *Lagged success rates* were used in this study as a proxy variable of a learning mechanism and hence reducing the risk of exploration. In all the models estimated in this chapter, lagged *success rate* has a positive and significant coefficient. This implies that firms reduce uncertainty by learning from the success history of previous exploratory wells drilled in the specified area of exploration.

The other stylized facts in the economics of non-renewable resources are clustering of deposits and depletion effect of reserves. On the one hand, as intensive exploration is carried out in a region, firms would tend to put more exploration effort based on the fact that deposits tend to cluster. This implies that the probability of finding an oil or gas deposit is higher in the area with more clustering. On the other hand, as more exploration and development is carried out in the region, the number of unexplored areas decreases and depletion effect will start to take place. According to Cairns (1990), this phenomenon is explained as the common access problem of exploration land, which can lead to a fishery-type problem of excess activity. Clustering of deposits and depletion effects are addressed in the empirical model by including the cumulative number of wells and square of cumulative wells as explanatory variables.

4.5 Summary

The major findings in this chapter can be summarized as follows. First, a comparison between the spatial and non-spatial models shows that the spatial model performs better than the non-spatial model. Based on this comparison all regional models are estimated taking into account spatial autocorrelation in the data. Second, no major differences were observed between the spatial lag and spatial error models in terms of the significance and magnitude of the coefficients. The spatial error model performed better in terms of higher R-squared values but the spatial lag model was chosen for prediction purposes on the basis of theoretical grounds. Third, significant differences were observed among the coefficients estimated for Alberta and the three regions within Alberta. Comparisons among the three regions show that there is some similarity among coefficient estimates. The price elasticity of drilling for the Northern and Plains regions are also similar. However, price elasticity for the Foothills region behaves differently. For example, the average price elasticity of drilling is 0.45 and 0.47 for the Northern and Plains regions and 0.97 for the Foothills region. Fourth, evidence of clustering of deposits and the depletion effect of reserves are observed in all models. The inclusion of lagged *success rate* as one of the explanatory variables is evidence of exploration learning effects. The positive and significant coefficient of this variable shows that firms gain information from previous period discovery and incorporate this information on the next round of exploration. This is one way of reducing uncertainty in the exploration process.

The models developed and estimated in chapters two and four are used in the next chapters. In chapter 5, the model is used to forecast drilling activities for the three regions in Alberta. The spatial lag model is used to forecast exploration activities and the results are compared with non-spatial models. In chapter 6, the model is applied to examine the response of the energy sector drilling activities to the anticipation of new regulations pertaining to the protection of wildlife habitat. In other words, the spatial lag model is applied to investigate exploration activities in caribou habitat before and after precautionary information about the status of woodland caribou is released. Specifically, the spatial lag model is used in the context of multivariate regression model and the difference in difference model where caribou habitat is included as one of the explanatory variables. Specifications of these models are explained in chapter 6.

Chapter 5 Forecasting oil and gas exploration activities

5.1 Introduction

The aim of this chapter is to forecast oil and gas drilling activities on the landscape of Alberta. Based on the model estimated in chapter 4, spatial and temporal forecasts of drilling effort are made for the three regions. The forecast is made at a township level up to the year 2020. Forecasting of exploration activities is very important in order to understand the extent of future energy sector footprint on the landscape. This information should be useful to land use planning and management. On the one hand, it is widely known that Alberta is a busy landscape with different sectors engaged in land use developments in which the energy sector is one of the dominant players. On the other hand, habitat conservation for wildlife and biodiversity are immediate issues that need to be balanced with land use developments. To this end, understanding future energy sector exploration activities is an important input to the integrated land use and cumulative effects management initiatives[26].

The present study is also an integral part of the Boreal Ecology and Economics Synthesis Team (BEEST) project. The aim of this project is to improve our understanding of the spatial economic behavior of different sectors in Alberta such as forestry, wildlife, and the energy sector. An example from the BEEST project that incorporates the energy sector is the forest bird abundance model (Hauer et al 2007). The model is estimated as expected bird counts for a given location using the factors that affect bird abundance. These factors include different types of forest stands, geographic variations in bird population, and industrial activities such as forest and energy sector. The energy sector is represented in the model as the number of oil and gas wells. The results from the present study, specifically forecasts of drilling effort through time and space, are important inputs in to this modeling and tradeoff analysis exercise.

---

[26] The Government of Alberta is taking an initiative to develop and implement land use planning in order to address a wide range of land management issues. More information is found at www.landuse.gov.ab.ca.

Forecasting using spatial panel data has it own difficulties. Statistical techniques that are commonly used in time-series analysis are not easily generalized and applied to spatial panel data due to spatial dependence in the model. Neglecting such dependence might result in sub-optimal forecasts. Forecasting using spatial panel data models is not very common in the economics literature. Some of the few studies include Longhi and Nijkamp (2007) and Baltagi and Li (2004, 2006) [27]. The main purpose of the study by Longhi and Nijkamp (2007) was to assess whether spatial econometric techniques such as the spatial lag and spatial error models would improve forecasting performance relative to the non-spatial model. The authors use data from 326 West German regions for the period of 1987 – 2002. Their results suggest that imposing a spatial structure improves the forecasting performance of non-spatial models. Using a similar setting, Baltagi and Li (2004, 2006) use spatial panel data models to predict cigarette consumption and the demand for liquor in the US. The demand equation for liquor was based on a panel of 43 states over the period of 1965–1994. They took spatial autocorrelation due to neighboring states and individual heterogeneity across states into account. Using a root mean squared error criterion, they found that predictions that take into account spatial correlation and heterogeneity across the states perform better than non-spatial models. Similar results were obtained for the cigarette consumption model. The present study follows similar statistical procedures to forecast drilling effort on the landscape of Alberta.

---

[27] Recent unpublished studies on this area include Kholodilin et al (2007) and Patuelli et al (2006). These studies deal with forecasting regional GDP and unemployment in German states respectively.

5.2 The forecasting model and data

5.2.1 The process of forecasting

Forecasting drilling activities for each township in Alberta is performed based on the parameter values estimated in chapter 4 and based on time paths for the explanatory variables. These variables are specified as projections of a historical series or assumed trajectories of the future behavior of the series. Descriptions of each explanatory variable are given in the next section. For prediction purposes the spatial lag model is used based on the theoretical discussions in chapters 2 and 4. The prediction equation can be written as follows:

$$\hat{w}_{it} = F\left(W\hat{w}_{i(t-1)}, A_{i(t-1)}, A^{2}{}_{i(t-1)}, \hat{S}_{i(t-1)}, P_t, T_t, CU_t, D_t / \hat{\delta}_w, \hat{\beta}_k\right) \qquad 5.1$$

where $\hat{w}_{it}$ refers to predicted values of drilling effort in township i at time t. W is the weight matrix as defined in chapter 2 and $\hat{\delta}_w$ is the coefficient of the spatial lag variable. According to the spatial lag model, the spatial lag of the dependent variable is included as one of the explanatory variables. However, for forecasting purposes, one period lag of the spatial-lag variable is used as a proxy for the current spatial-lag variable since the number of wells for the current period is yet to be forecasted. Hence, $W\hat{w}_{i(t-1)}$ is used on the right hand side of equation 5.1 instead of $W\hat{w}_{it}$. $\hat{\beta}_k$ s are estimated coefficients for the k explanatory variables. $\hat{S}_{i(t-1)}$ refers to the estimated *lagged success rate* in township *i* in period *t-1*. $A_{i(t-1)}$ and $A^2{}_{i(t-1)}$ are lagged cumulative wells and its square respectively. P_t, T_t ,CU_t, and D_t represent *price, technology, capacity utilization*, and a dummy variable to capture *3D seismic survey* respectively. Instead of assuming a fixed success rate for the forecast period, the *success rate* in equation 5.1 is endogenously determined by an expression including the cumulative number of wells and its squared valued. The equation for the *success rate* can be written as:

$$\hat{S}_{i(t-1)} = F(A_{i(t-2)}, A^2{}_{i(t-2)} / \hat{\alpha}) \qquad 5.2$$

were $\hat{\alpha}$ refers to a vector of the estimated coefficients for the cumulative number of wells and its square. During the forecasting process equation 5.2 is one of the explanatory variables for equation 5.1. This is a two stage process where equation 5.2 is estimated first and the values of the estimated equation are incorporated into equation 5.1.

Forecasting is implemented using a step by step process. First, initial values of the explanatory variables for the forecasting period (t = 2005) are obtained and then the predicted values for the forecasting year are calculated. For the second period forecast (t = 2006), predicted values from the previous year forecasts are used to update the cumulative number of wells and squared cumulative wells. Simultaneously, the *success rate* variable for the second forecasting period is updated using equation 5.2. The projected series for the other explanatory variables are then incorporated to calculate the second period forecast. For each forecasting period, lagged values of the spatial-lag variable are calculated using GIS software[28] and exported to the forecasting worksheet. The process continues one period at a time for the entire forecasting period. Data on the projected series of the explanatory variables are discussed in the next section.

5.2.2 Projected data for explanatory variables

The process of forecasting requires data on the explanatory variables for the specified forecasting period. The process of acquiring these variables is discussed below:

Spatial lag of wells: Data on the current period of the spatial lag variable are yet to be forecasted. Hence, one period lagged values of the spatial-lag variable are used as a proxy for the current period.

Cumulative wells: Projections of cumulative wells are made one year at a time. For example, forecasted values of incremental wells from $T = \tau$ will be added to the cumulative wells for $T = \tau - 1$ in order to calculate the cumulative wells for $T = \tau$. This variable and its squared value are then used to forecast the number of wells for $T = \tau + 1$.

---

[28] The software Geodà 0.9.5-i (2004) developed by Luc Anselin is used to calculate spatial lag values.

Success rate: Successful wells are not known beforehand because of the uncertainty in the exploration outcome. Hence, projections of *success rate* are endogenously determined based on equation 5.2. The outcome from this equation is incorporated into equation 5.1.

Price index of oil and gas: Forecasts for the price of oil and gas are based on the Alberta Energy and Utilities Board (EUB) forecasts. The EUB forecasted that the price of oil will average between US$55 and US$60 per barrel for 2005 and 2006, before it stabilizes at US$55 by 2009 (EUB, 2005). The EUB forecasts of Alberta's wellhead price for oil are given in Figure 5.1 below. Historical and EUB forecasts of natural gas prices at the plant gate from 1995 to 2020 are given in Figure 5.2. Forecasts of the price index for oil and gas are then calculated using forecasts of oil and gas prices. Analysis of a 10 and 20 percent increase in price are also included in the forecasting exercise.

Technical change and capacity utilization: A trend variable is used as a proxy for technical change. The variable is specified as t = 1 for the initial study period (year 1980) and t = 25 for the year 2004. For forecasting purposes the natural logarithm of time series values of the trend variable are used to represent technical change throughout the projection period. For example, for the first forecasting period, t = 26 is used to represent a trend for the year 2005 and so on. As discussed in chapter 3, a *capacity utilization* rate is calculated as the number of active rigs divided by the total number of available rigs in a given year. The average *capacity utilization* rate is used for the entire forecasting period. Having discussed the prediction model and the forecasted values of the explanatory variables, the next section discusses model validation.

Figure 5.1 Forecasts of average price of oil at Alberta wellhead (Source EUB 2005)[29]

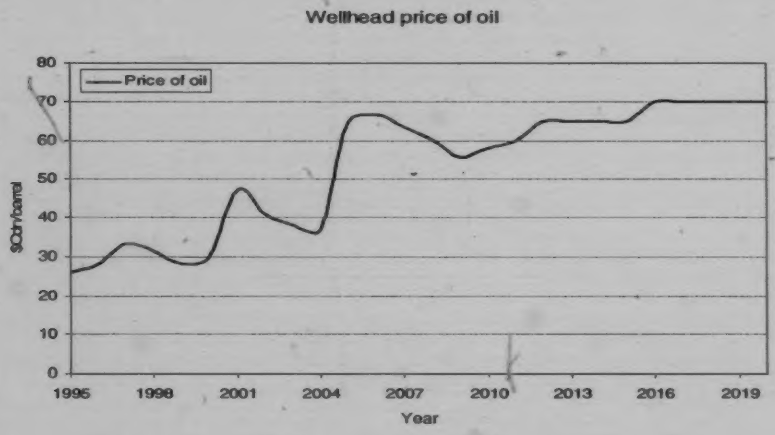
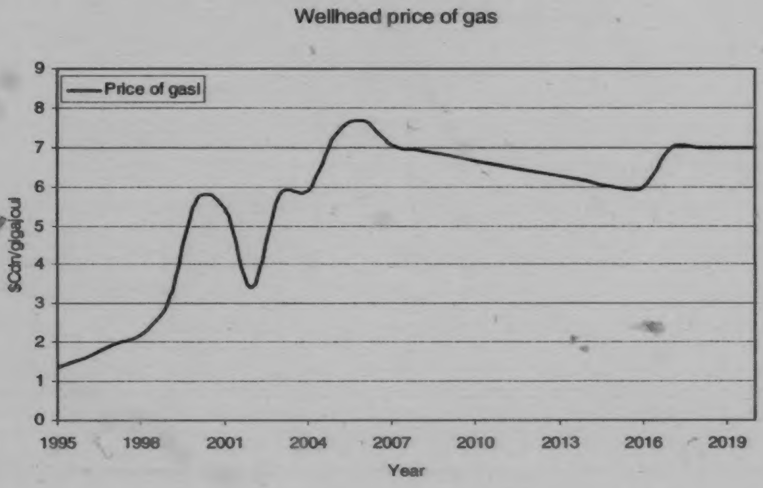


Figure 5.2 Forecasts of average price of natural gas at plant gate (Source EUB 2005)



---

[29] The price of oil or gas shows average value for the specified year. Even though current prices are higher the above forecasts show estimates using conservative price increases.

5.3 Model validation

5.3.1 The process of validation

Model validation is an important part of model development, estimation, and forecasting processes. The estimated model should be validated to ensure that the model meets its intended requirements in terms the methods employed and the results obtained. To this end, the model estimated in chapter 4 is re-estimated by holding back 20 percent of the data for validation purposes. In other words, data from 1980 - 2000 are used to re-estimate the model and based on the results; forecasts for the year 2001 – 2004 are obtained. These forecasts are then compared to the actual values using different statistical indicators. Unlike the common forecasting procedures in the time-series literature, in the present study, forecasting has to be made for each township per year. The statistical indicators are then calculated based on one year *ex post* forecasts. For example, the correlation coefficient for the *ex post* forecast of time t = 2001 is computed as the correlation between the actual number of wells drilled in township i in the year 2001 (w_{i2001}) and the predicted number of wells drilled in township i in the year 2001 ($\hat{w}_{i2001}$). The average correlation for the four year period is then calculated using the yearly correlation coefficients. The same procedure is applied to the other statistical indicators.

The statistical indicators used to compare the *ex post* forecasts with the actual forecasts are the correlation coefficient, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The following equations provide the formulae for each statistical indicator:

The correlation coefficient is given as $\rho = \frac{\sum_i (w_{it} - \overline{w}_{it})(\hat{w}_{it} - \overline{\hat{w}})}{(N-1)\sigma_w \sigma_{\hat{w}}}$, where w_{it} and $\hat{w}_{it}$ are the actual and predicted values of drilling densities in township i at time t respectively and the bar shows mean values. N refers to the number of townships. σ_w and $\sigma_{\hat{w}}$ refer to the standard deviations of the actual and predicted values. The correlation coefficient measures the statistical correlation between the predicted and the actual values. Mean absolute error (MAE) shows the average of the difference between predicted and actual values. $MAE = \frac{\sum_i |w_{it} - \hat{w}_{it})|}{N}$. In other words, MAE measures average prediction error. Mean squared error (MSE) is computed by taking the average of the squared differences between the actual and predicted values. $MSE = \frac{\sum_i |w_{it} - \hat{w}_{it})|^2}{N}$. The smaller the MSE, the closer the fit is to the data. The root mean-squared error (RMSE) is simply the square root of the mean-squared error. The RMSE is thus the distance, on average, of a data point from the predicted value to the actual value measured along a vertical line[30].

5.3.2 Coefficient estimates for the validation model

Two different versions of the models are estimated for validation purposes, the spatial and non-spatial models. Table 5.1 shows the results for the *success rate* equation for the three regions in Alberta. Results for the non-spatial panel data model are reported in column two and results for the spatial lag model that takes into account spatial and time-period fixed effects are given in column three of the table. Results from Table 5.1 show that the signs of the coefficients are consistent for the three regions. Cumulative wells are included in the success rate equation to capture reserves and infrastructure variables and the square of cumulative wells represents the depletion effect. The positive coefficient for cumulative wells confirms that success rates are higher in townships where firms have adequate information on the availability of reserves and the negative coefficient shows the depletion effect. The spatial model performs better than the non-spatial model. This is reflected in higher R-squared values for each region. The positive and significant coefficient for the spatial lag variable explains that information gained from neighboring townships would help firms to drill successful wells in a given township.

---

[30] The concept and definitions of these statistical indicators and their formulae are taken from Longhi and Nijkamp (2007).

Table 5.2 contains results of the drilling equation for the three regions in Alberta. Results from a non-spatial panel data model and a spatial lag model are reported under the specified columns for each region. The signs of most of the coefficients are as expected and they are consistent with the model estimated in chapter 4. The spatial lag model with an R-squared equal to 0.28 performs better than the non-spatial model (R-squared equal to 0.19).

Table 5.1 Results of the success rate equation: model validation sample (1980 – 2000)

	Non-spatial model	Spatial /time effects Model
Northern Region		
Variable	Coefficient (t-stat)	Coefficient (t-stat)
Constant	0.046 (3.4)	-
Cum. Wells	0.600 (43.8)	0.456 (47.5)
Cum. Wells Squared	-0.120 (-22.8)	-0.070 (-30.4)
*W*dependent variable*	-	0.036 (6.7)
N	67,870	67,870
R-squared	0.14	0.27
Plains Region		
Constant	0.150 (2.9)	-
Cum. Wells	0.210 (46.5)	0.180 (51.2)
Cum. Wells Squared	-0.012 (-33.9)	-0.010 (-28.4)
*W*dependent variable*	-	0.043 (6.0)
N	44,792	44,792
R-squared	0.11	0.21
Foothills Region		
Constant	-0.002 (3.2)	-
Cum. Wells	1.577 (28.7)	3.049 (34.6)
Cum. Wells Squared	-1.264 (-22.8)	-2.969 (-25.6)
*W*dependent variable*	-	0.017 (8.4)
N	11,946	11,946
R-squared	0.19	0.28

Note: t-statistics in parenthesis.

Table 5.2 Results of the drilling effort equation: model validation sample (1980 – 2000)

	Northern		Plains		Foothills	
	Non-Spatial	Spatial /time effects	Non-Spatial	Spatial /time effects	Non-Spatial	Spatial /time effects
Variable	Coef	Coef	Coef	Coef	Coef	Coef
Constant	-0.010 (-9.8)	-	-0.040 (-7.3)	-	-0.002 (-3.2)	-
Cum. Wells	0.050 (54.3)	0.014 (7.7)	0.040 (51.3)	0.040 (41.1)	0.050 (24.0)	0.060 (11.1)
Cum. Wells Squared	-0.002 (-9.3)	-0.003 (-7.9)	-0.001 (-4.6)	-0.001 (-13.6)	-0.040 (-11.3)	-0.070 (-11.8)
Lag. Success Rate	0.020 (36.2)	0.020 (32.8)	0.040 (31.4)	0.030 (28.2)	0.010 (27.1)	0.010 (22.0)
Lag Price-1 Period	0.003 (4.9)	0.005 (12.4)	0.008 (3.4)	-0.006 (-4.0)	0.002 (5.1)	0.001 (2.2)
Lag Price -2 Period	0.003 (2.6)	0.002 (4.3)	-0.005 (-1.6)	0.007 (3.5)	0.001 (1.7)	-0.001 (-3.3)
Lag Price-3 Period	-0.002 (-1.5)	-.000 (-5.1)	0.008 (2.2)	0.028 (23.1)	-0.001 (-2.1)	0.003 (6.7)
Capacity Utilization	0.020 (13.8)	0.020 (24.3)	0.070 (12.0)	0.008 (7.3)	0.003 (3.5)	0.002 (8.1)
Time Trend	0.000 (0.9)	0.002 (7.2)	-0.003 (-3.8)	-0.002 (-3.4)	0.000 (-0.9)	-0.000 (-1.1)
3D Seismic Dummy	-0.001 (-2.6)	-0.001 (-2.3)	0.000 (0.4)	-0.001 (-0.8)	0.001 (1.1)	0.001 (2.2)
*W*Dep. Variable*	-	0.040 (6.7)	-	0.050 (28.2)	-	0.030 (2.7)
N	67,870	67,870	44,792	44,792	11,946	11,946
R-Squared	0.21	0.24	0.22	0.27	0.28	0.30
Adj.R-Squared	0.17	0.21	0.21	0.23	0.24	0.26
Log-Likelihood	130767.7	132245.5	47138.5	47936.7	44593.9	45538.5

Note: t-statistics in parentheses

5.3.3 Comparing ex post vs. actual forecasts

The results reported in Tables 5.1 and 5.2 are used to calculate the predicted values of drilling effort for the models specified in the tables. Graphical representations of the average values for the actual and predicted values of the Northern, Plains, and Foothills regions are given in Figures 5.3, 5.4, and 5.5 respectively. In line with these figures, Table 5.3 summarizes results of the statistical indicators discussed in sub-section 5.3.1.

Figure 5.3 Northern region: actual versus ex post forecasts

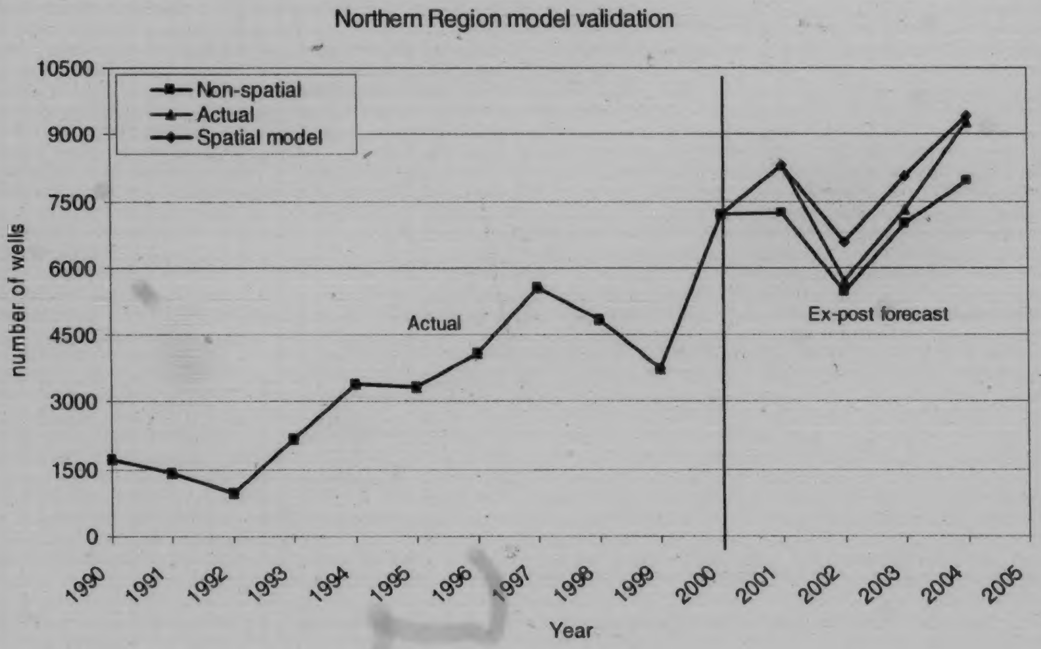


Figure 5.4 Plains region: actual versus ex post forecasts

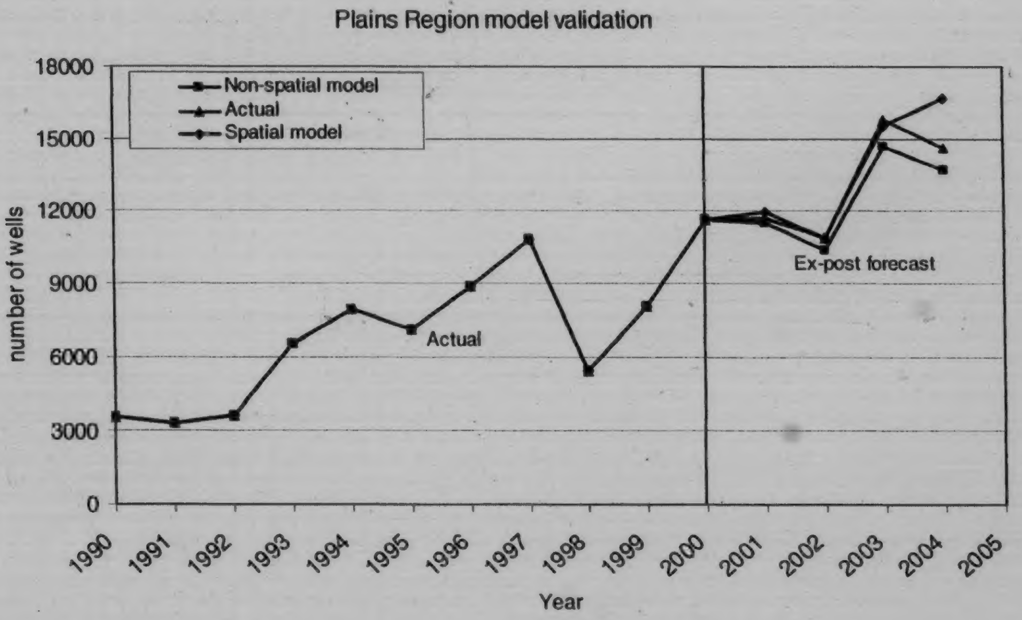
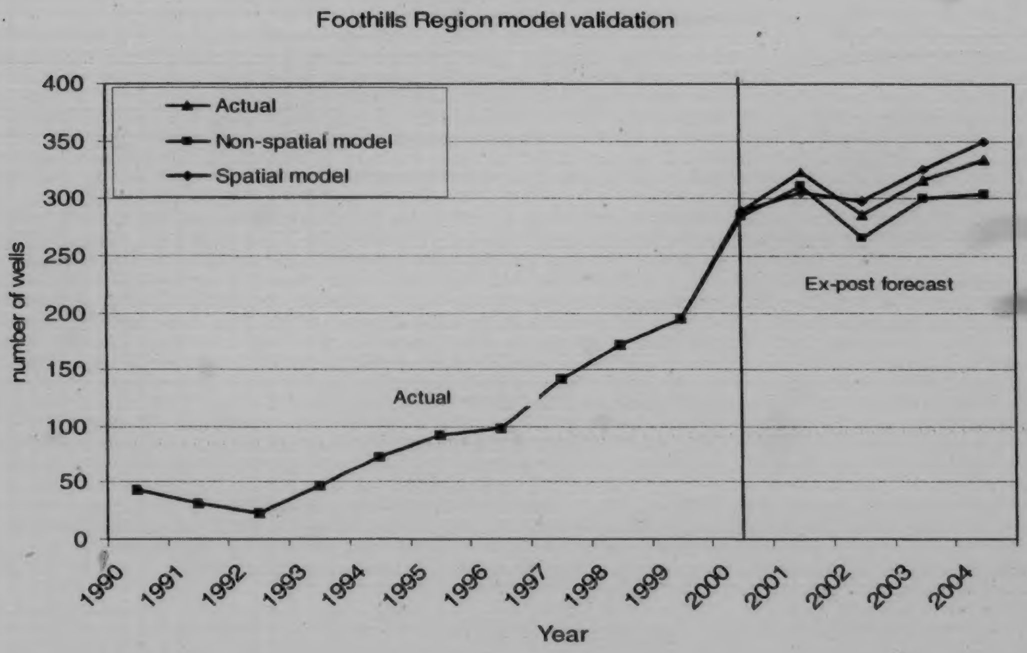


Figure 5.5 Foothills region: actual versus ex post forecasts



Figures 5.3 to 5.5 illustrate the differences in forecasts when the spatial and non-spatial models are used to validate the forecasting model. One of the common features among the three figures is that on average the non-spatial model tends to underestimate the actual values.

Table 5.3 Comparison of actual vs. ex post forecasts using statistical indicators

Northern region

Measures	Non-spatial	Spatial model
MAE	-0.002	-0.002
MSE	0.003	0.002
RMSE	0.058	0.051
Rho (ρ)	0.575	0.601

Plains region

Measures	Non-spatial	Spatial model
MAE	0.005	0.002
MSE	0.000	0.000
RMSE	0.001	0.001
Rho (ρ)	0.424	0.520

Foothills Region

Measures	Non-spatial	Pooled spatial
MAE	0.0004	0.0002
MSE	0.000	0.000
RMSE	0.000	0.000
Rho (ρ)	0.501	0.563

Table 5.3 shows that the correlation coefficients between the actual and the forecasted values are higher for the spatial model in all regions. For example, the correlation between the actual and forecasted values is 0.57 for the non-spatial model and 0.60 for

the spatial model. Comparisons between the mean squared errors (MSE) show that the spatial model has a lower MSE than the non-spatial model for the three regions. How significant are these differences statistically? The U-Theil inequality coefficient is often cited in the literature as a method of comparing MSE of models (Swanson and White 1997). This coefficient is computed as the ratio between the MSE of the first model to the MSE of the second model. The proposed model outperforms the base model when the U-Theil coefficient is lower than one (Swanson and White 1997). Assuming that the non-spatial model is the base model and the spatial model is the proposed model, the U-Theil coefficients are 0.79, 0.88, and 0.95 for the Northern, Plains, and Foothills regions respectively. In all cases, the spatial model outperforms the non-spatial model. For the Foothills region the coefficient is closer to one. This could indicate that within the Foothills regions the spatial and aspatial models have similar forecasting performance. In sum, the statistical indicators show that the spatial model performs better than the non-spatial model.

5.4 Model results and forecasts

Based on the model validation results, the spatial lag model that takes into account time period and spatial fixed effects is used to forecast drilling effort. Coefficient estimates of the spatial lag model for drilling effort and *success rate* equations are reported in Tables 5.4 and 5.5 respectively. Results in Table 5.4 are obtained from spatial lag regional models estimated in chapter 4. Hence, we do not explain the results for each variable here. Results for the success rate equation in Table 5.5 are similar to the validation model reported in Table 5.1. The only difference is the sample size where the new results are based on the entire study period (1980 – 2004).

Table 5.4 Results of the drilling effort equation for the three regions: (1980 – 2004)

	Northern	Plains	Foothills
Variable	Coef	Coef	Coef
Constant	-	-	-
Cum. Wells	0.050	0.030	0.070
	(42.6)	(17.2)	(18.5)
Cum. Wells Squared	-0.005	-0.003	-0.06
	(-21.1)	(-26.2)	(-13.1)
Lag. Success Rate	0.020	0.040	0.010
	(37.2)	(30.8)	(24.7)
Lag Price-1 Period	0.003	0.010	0.001
	(9.5)	(11.5)	(8.2)
Lag Price -2 Period	0.001	-0.000	0.000
	(4.1)	(-0.2)	(1.8)
Lag Price-3 Period	-.000	0.009	-.0001
	(-0.9)	(8.4)	(-0.6)
Capacity Utilization	0.020	0.080	0.003
	(21.5)	(26.0)	(5.1)
Time Trend	-0.000	-0.005	-0.001
	(-0.4)	(-0.7)	(-2.1)
3D Seismic Dummy	-0.001	0.003	0.000
	(-3.9)	(2.9)	(1.3)
*W*Dep. Variable*	0.040	0.060	0.010
	(9.1)	(32.1)	(7.2)
N	77125	50900	13575
R-Squared	0.25	0.30	0.30
Log-Likelihood	138270.7	50774.4	45113.0

Note: t-statistics in parentheses

Table 5.5 Results of the success rate equation for the three regions: (1980 – 2004)

	Model estimated using spatial lag time and spatial fixed effects		
	Northern	Plains	Foothills
Variable			
Cum. Wells	0.393 (54.5)	0.213 (63.2)	2.718 (46.6)
Cum. Wells Squared	-0.053 (-34.5)	-0.011 (-33.6)	-2.264 (29.0)
*W*dependent variable*	0.035 (7.0)	0.033 (5.7)	0.013 (8.3)
N	77,125	50,900	15,325
R-squared	0.27	0.24	0.30

Note: t-statistics in parentheses

Forecasts of drilling effort for the years 2005 to 2020 are made using equations 5.1 and 5.2. Coefficient estimates reported in Tables 5.4 and 5.5 are substituted into these equations to derive forecasted values. A summary of the results for each region are given in Figures 5.6 to 5.8. Three forecasts are obtained for each region. The first series uses the base model that takes into account the spatial dependence variable. Four more forecasts were obtained by assuming an increase of oil and gas prices by 10, 20, 50, and 100 percent keeping the other independent variables constant. This was done to analyze the sensitivity of drilling effort to different ranges of price changes at a regional level. The results show that drilling in the Foothills region changes more substantially when price is increased by the above specified percents compared to the other two regions. For example, a 10 percent increase in price results in a 15, 11, and 29 percent increase in drilling effort in the Northern, Plains, and Foothills regions respectively. A 20 percent increase in price on average results in a 22, 17, and 32 percent increase in the Northern, Plains, and Foothills regions respectively.

Figure 5.6 Summary of forecasts for the northern region

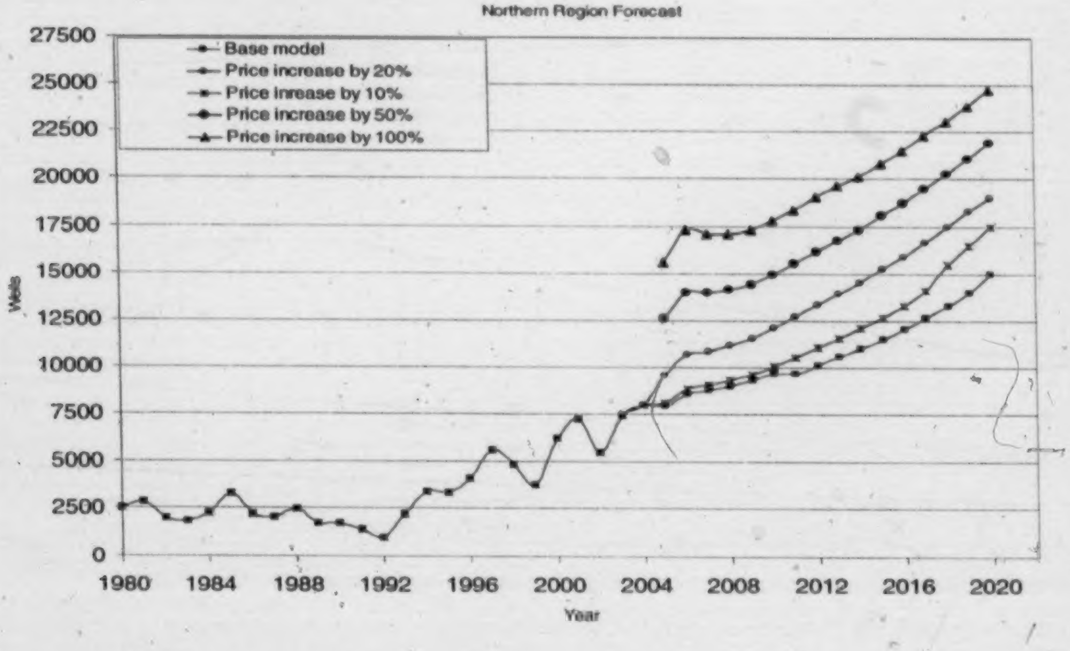


Figure 5.7 Summary of forecasts for the plains region

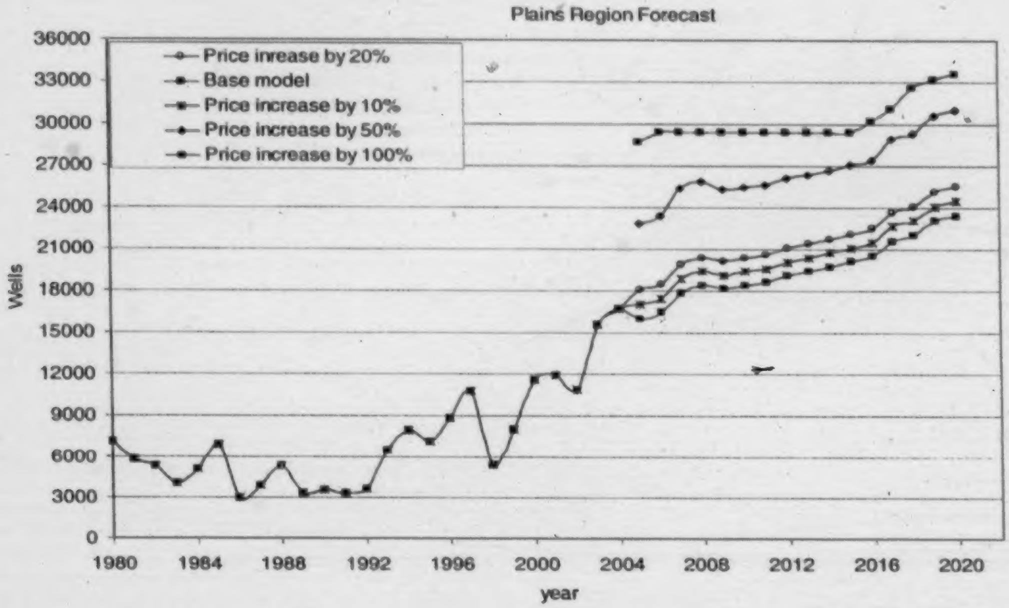
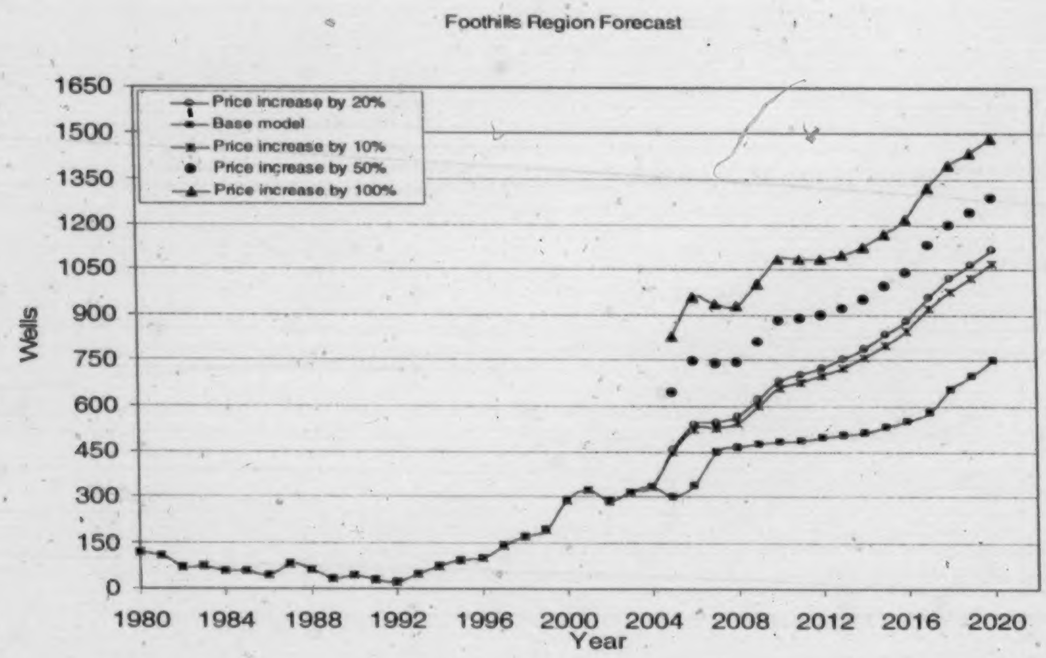


Figure 5.8 Summary of forecasts for the foothills region



Actual drilling in 2004 is given in Figure 5.9 and maps of forecasts of drilling effort for selected years are given in Figures 5.10 to 5.13. These maps are drawn based on the results from the base spatial lag model that takes time period and spatial fixed effects into account. Individual observations for each township per year are based on the forecasts made at a regional level. A closer look at the maps shows clustering of exploration activities in the southern and north-east parts of the province. The southern part includes extensive natural gas wells in the Medicine Hat region and the north-eastern part includes exploration activities in the Fort McMurray region. More clusters of exploration activities are observed beyond 2010 along the eastern, west central and central parts of Alberta.

Figure 5.9 Map of actual wells for the year 2004.

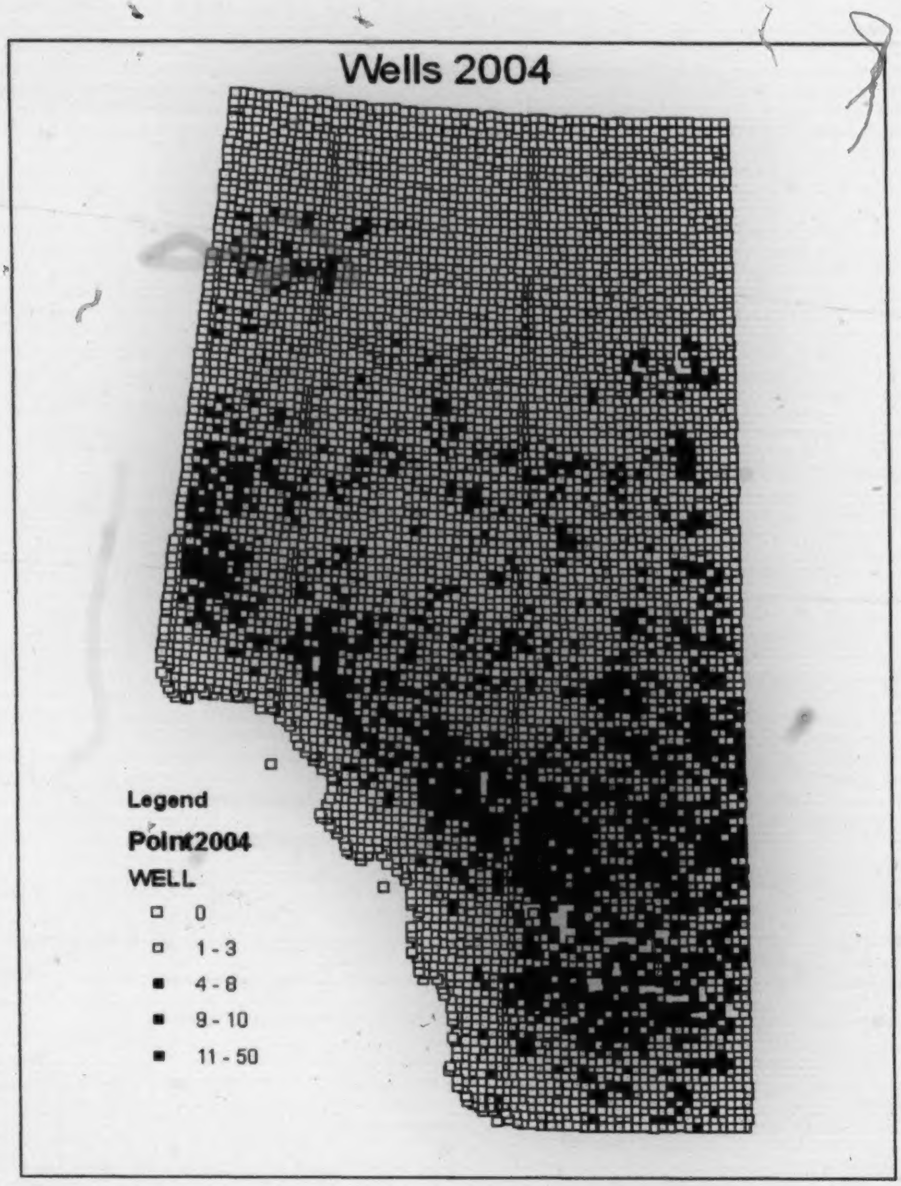

Wells 2004
Legend
Point2004
WELL
0
1 - 3
4 - 8
9 - 10
11 - 50

Figure 5.10 Map of forecasted wells for the year 2005.

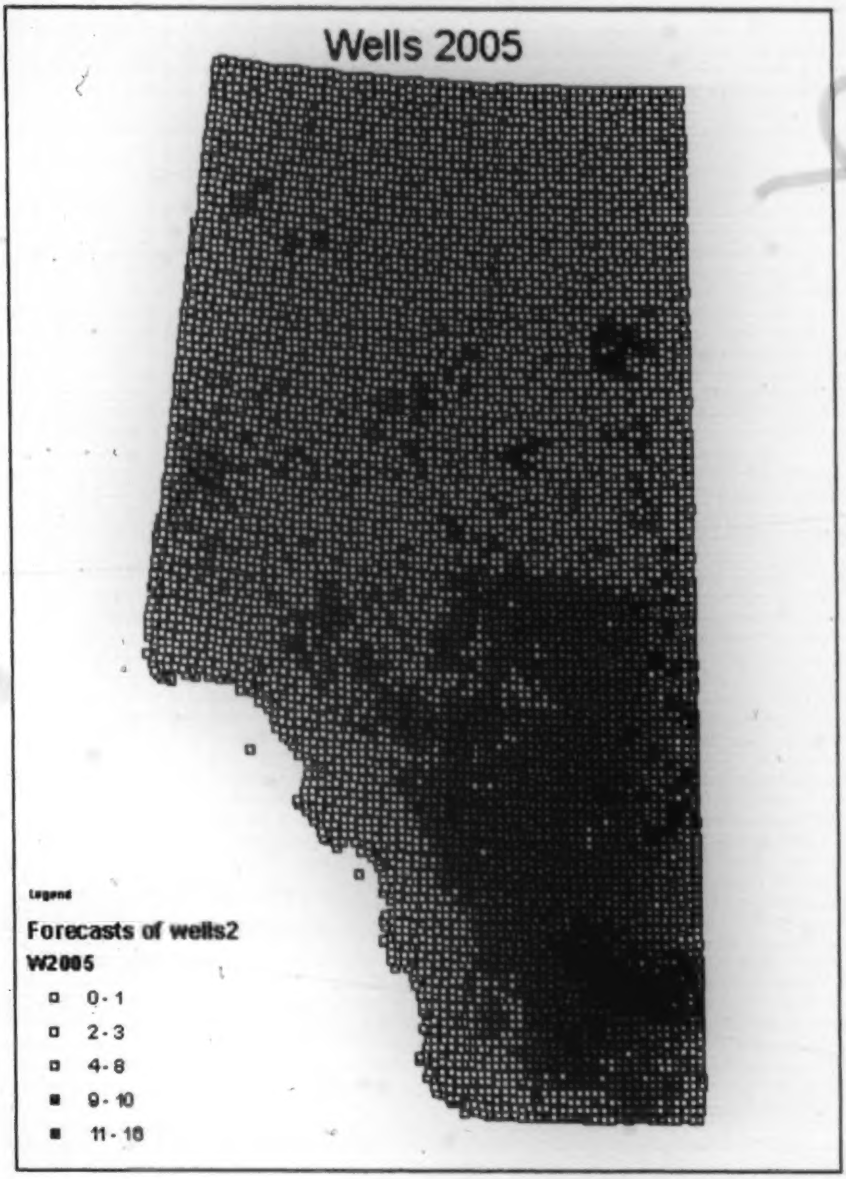

Wells 2005
Legend
Forecasts of wells2
W2005
0 - 1
2 - 3
4 - 8
9 - 10
11 - 18

Figure 5.11 Map of forecasted wells for the year 2010.

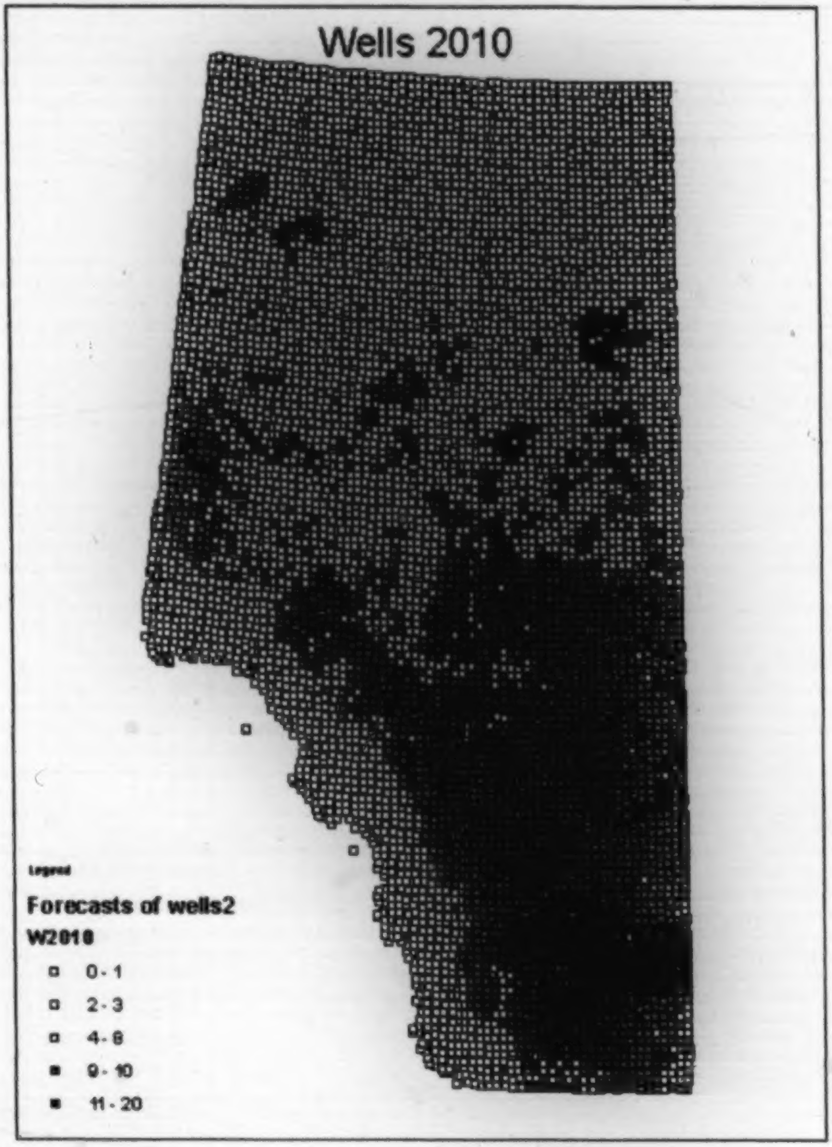

Wells 2010
Legend
Forecasts of wells2
W2010
0 - 1
2 - 3
4 - 8
9 - 10
11 - 20

Figure 5.12 Map of forecasted wells for the year 2015.

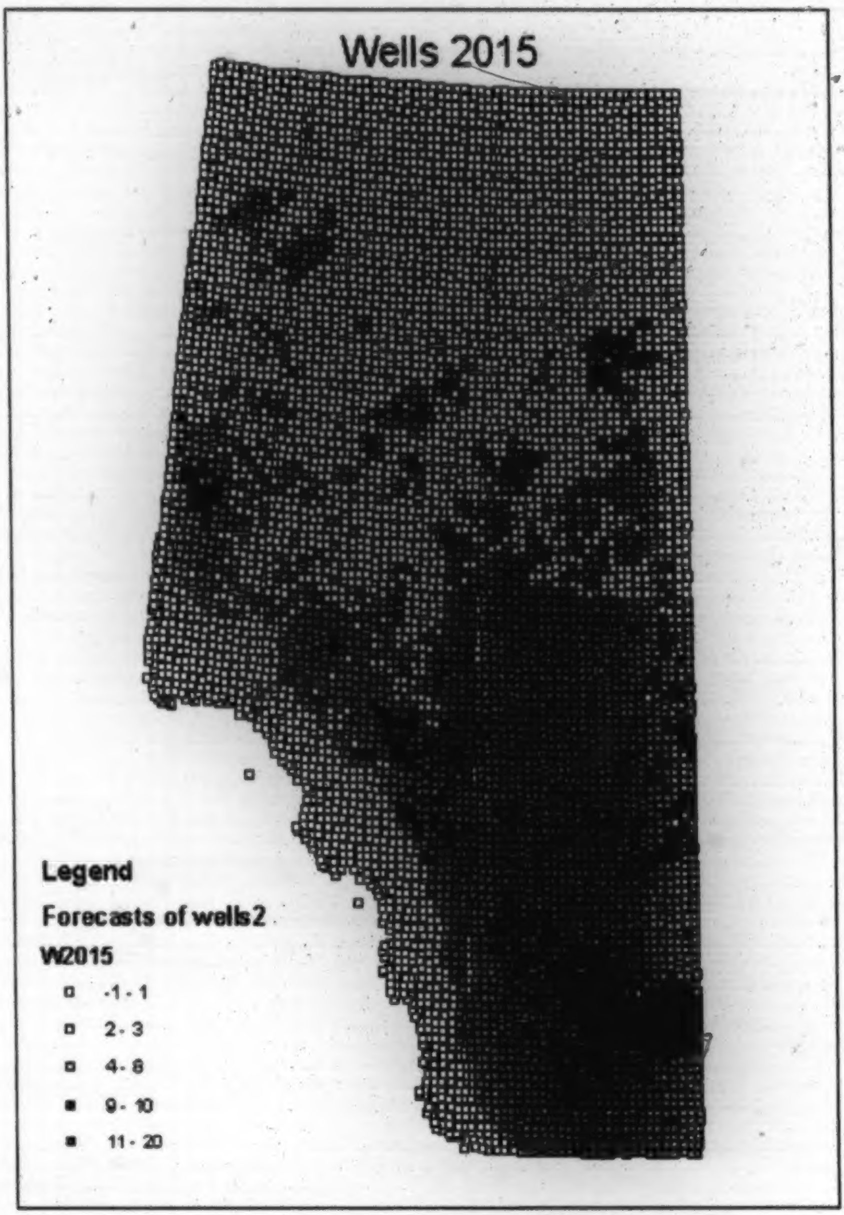
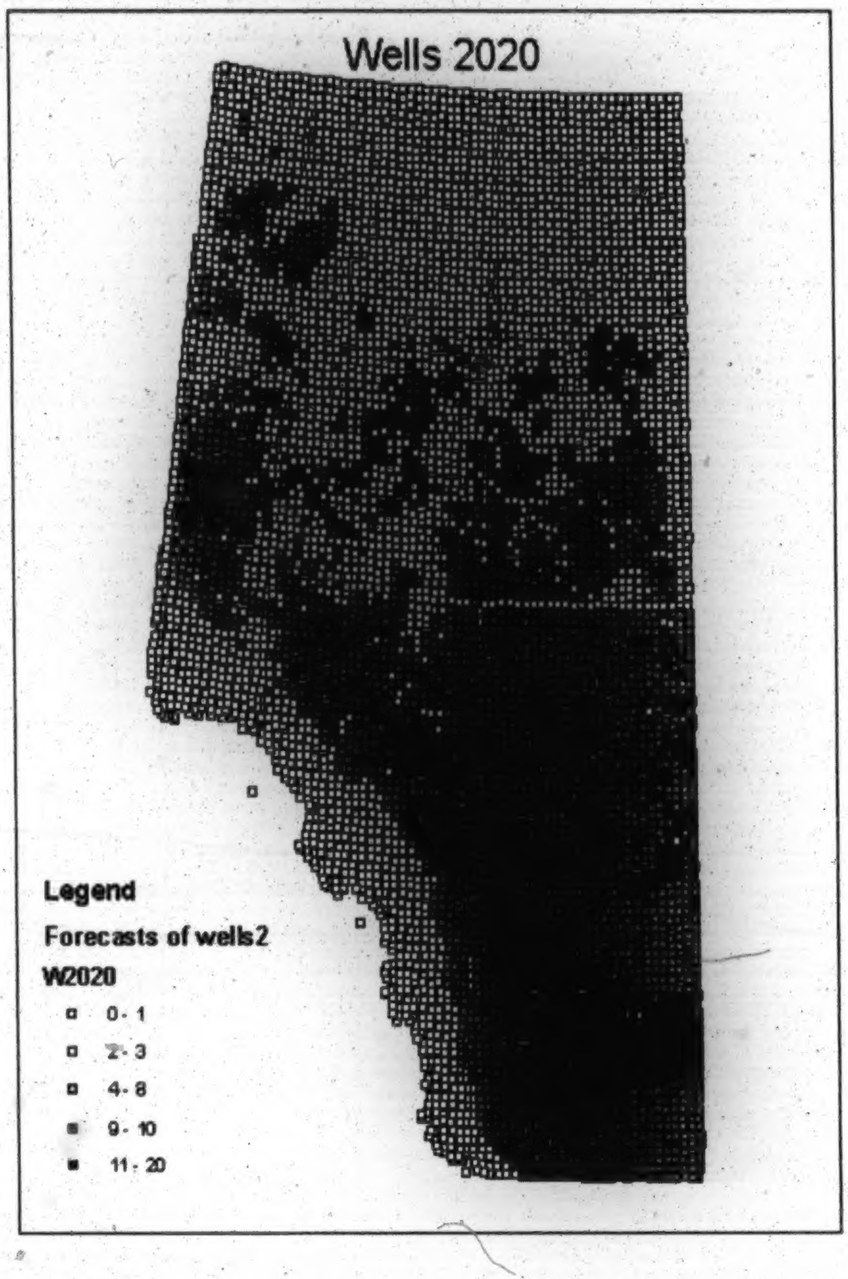

Wells 2015
Legend
Forecasts of wells2
W2015
-1 - 1
2 - 3
4 - 8
9 - 10
11 - 20

Figure 5.13 Map of forecasted wells for the year 2020.


Wells 2020
Legend
Forecasts of wells2
W2020
0 - 1
2 - 3
4 - 8
9 - 10
11 - 20

5.5 An application of the forecasting model

Forecasts of oil and gas wells at a spatial and temporal level have many applications. Projected exploration activities on the landscape of Alberta could be used to analyze the future rate of caribou population increase (λ) based on the model developed by the Alberta Woodland Caribou Recovery Team (2005). This model explains the relationship between caribou population trend and functional habitat loss resulting from anthropogenic activities. Given that oil and gas activity is one of the major anthropogenic disturbance in Alberta, this relationship can be used to evaluate the impact of human disturbance (drilling of oil and gas wells) in relation to the goals for the rate caribou population increase (λ).

Another application of the forecasting model is related to the forest bird abundance model (Hauer et al 2007). The bird model is estimated as expected bird counts for a given location using the equation: $b_{it} = f(X_i, G_i, Z_{it}, E_{it} : \beta)$; where b_{it} refers to bird counts at location *i* in period *t*, X captures different types of forest stands, G indicates geographic variations in bird population, Z are a set of other covariates and E is the energy sector variable expressed as number of wells. Given a significant energy sector activity, the results from the present study, specifically forecasts of drilling effort through time and space, are an important inputs in to this modeling and tradeoff analysis.

The study by Hauer et al. (2007) show that the number of oil and gas wells present in a given township has significant effect on several bird species. An example of the White-

Throated Sparrow (WTSP) bird model is taken as a case study to analyze the response of oil and gas drilling activities to bird counts in a given location. The coefficient of the energy sector variable for the WTSP bird model is used to forecast bird counts for the years 2010, 2015 and 2020. The analysis includes energy sector activities in the northern part of Alberta. Maps of forecasts of the bird counts are presented in Figures 5.14 to 5.16. The results displayed in these figures show that bird counts in a given township decrease as drilling activities increase through time.

Figure 5.14 Map of forecasted WTSP bird counts for the year 2010.

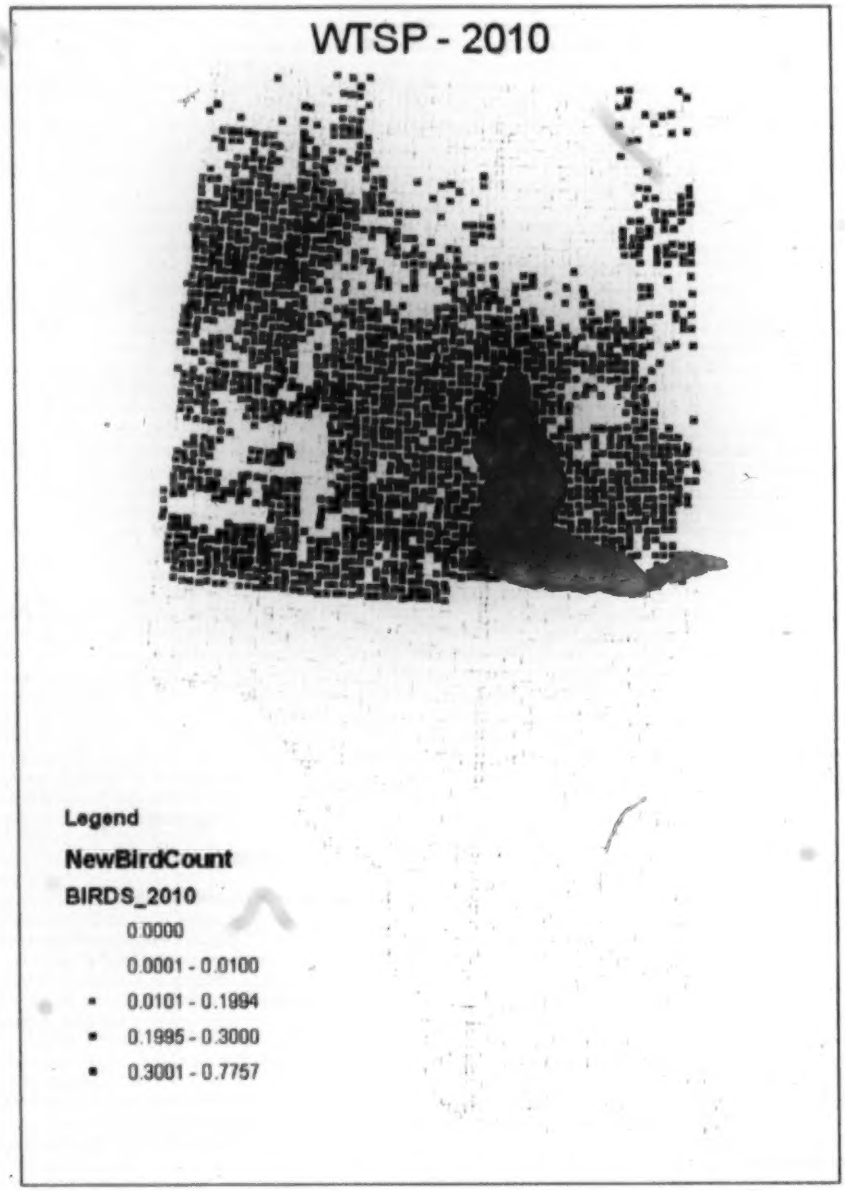

WTSP - 2010
Legend
NewBirdCount
BIRDS_2010
0.0000
0.0001 - 0.0100
▪ 0.0101 - 0.1994
▪ 0.1995 - 0.3000
▪ 0.3001 - 0.7757

Figure 5.15 Map of forecasted WTSP bird counts for the year 2015.

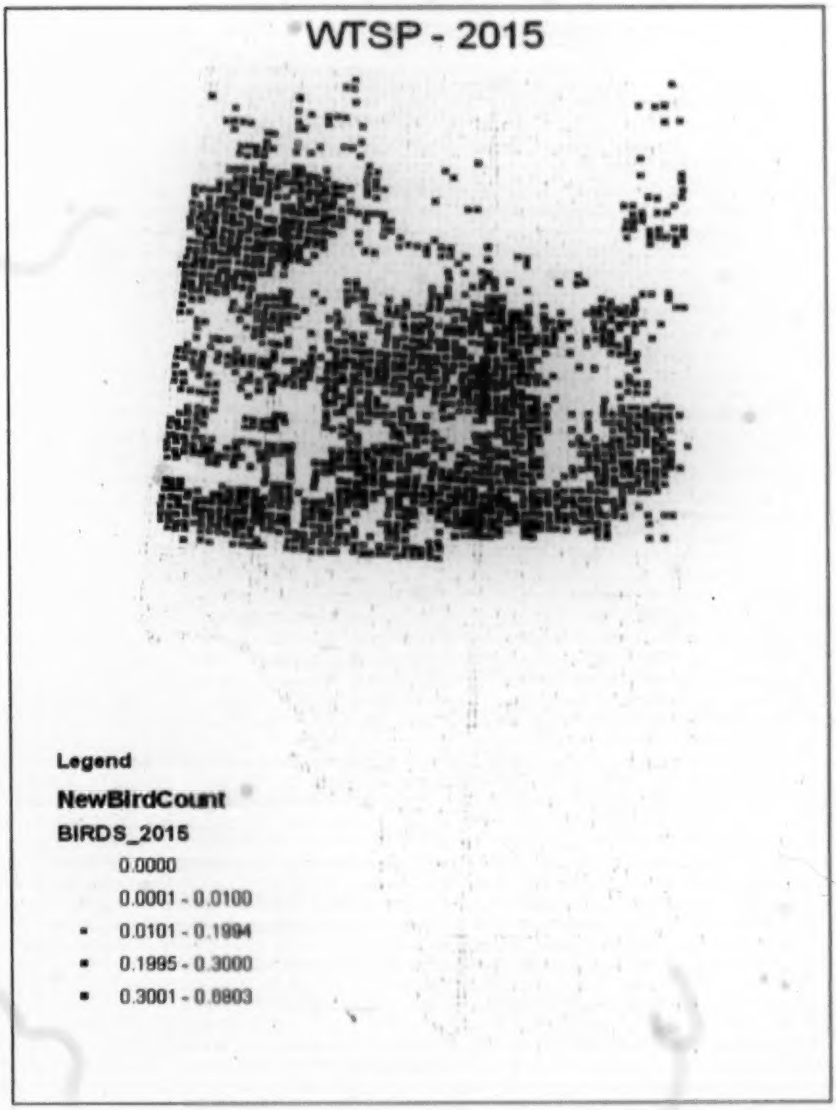
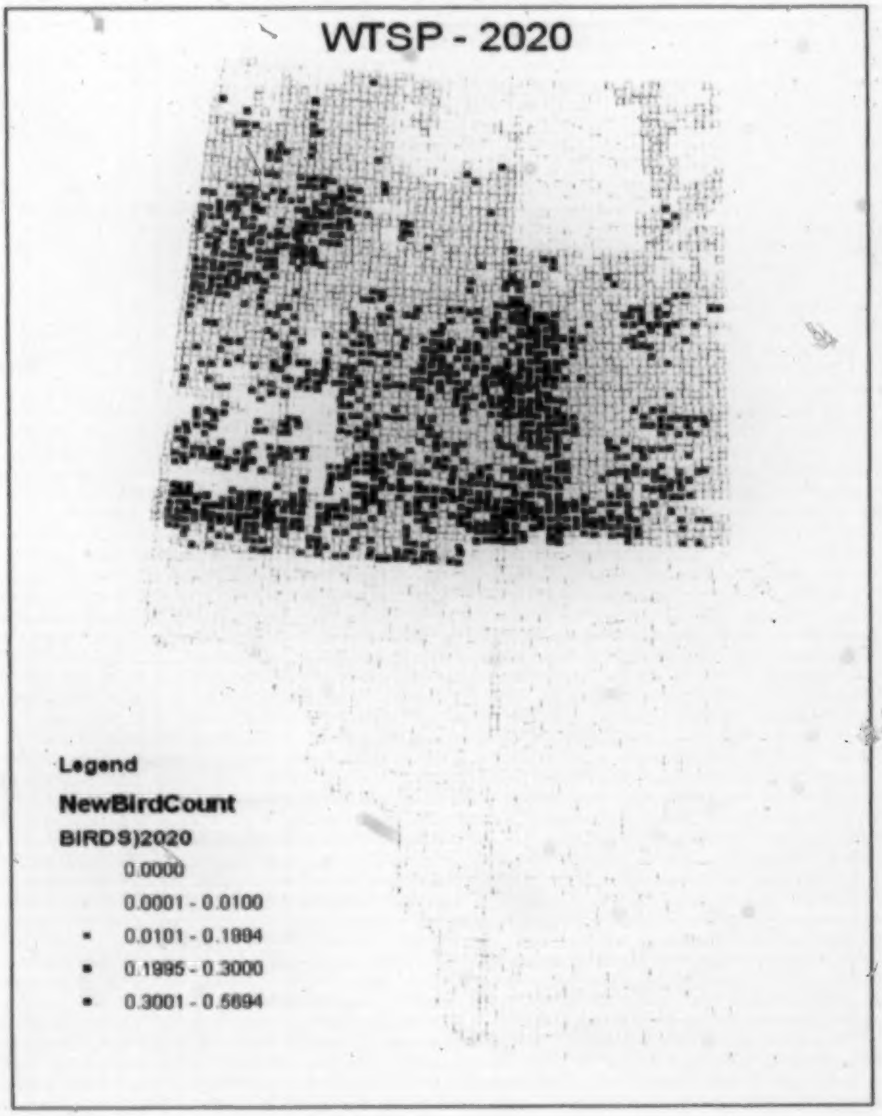

WTSP - 2015
Legend
NewBirdCount
BIRDS_2015
0.0000
0.0001 - 0.0100
0.0101 - 0.1994
0.1995 - 0.3000
0.3001 - 0.8803

Figure 5.16 Map of forecasted WTSP bird counts for the year 2020.


WTSP - 2020
Legend
NewBirdCount
BIRDS)2020
0.0000
0.0001 - 0.0100
▪ 0.0101 - 0.1994
▪ 0.1995 - 0.3000
▪ 0.3001 - 0.5694

5.6 Summary and discussion of the results

Forecasts of oil and gas drilling effort on a township basis were made for the three regions in Alberta. The forecasting performance of the model was validated using 80 percent of the data and holding the remaining 20 percent for comparison purposes. Ex-post forecasts were obtained using spatial lag and non-spatial models. A comparison between the actual and ex-post forecasts shows that the spatial lag model performs better than the non-spatial model for the three regions. Different statistical methods were used to compare these forecasts. For example, the correlation coefficient for the northern region was 0.60 for the spatial lag model compared with 0.57 for the non-spatial model. Moreover, the results show that the MAE, MSE, and RMSE of the actual versus the ex-post forecasts are all lower for the spatial models. Evaluation of the MSE using the U-Theil inequality coefficient criteria show that the spatial model performs better than the non-spatial model. Based on the model validation results the spatial lag model was chosen to obtain forecasts of drilling effort up to the year 2020. Sensitivity of drilling forecasts to price changes were examined assuming a 10 and 20 percent increase in prices. The results show that on average the percentage change in drilling, due to price change, is higher in the Foothills region than the other two regions.

Understanding forecasts of oil and gas drilling efforts on the landscape of Alberta are useful in many ways. For example, these forecasts are important in understanding ecological impacts as explained in the bird abundance model of the BEEST project (Hauer et al 2007). In a more general assessment of ecological impacts, forecasts of oil and gas drillings can be incorporated in to a model that provides a framework for integrated planning of industrial activities under ecological forest management (Schneider (2002). Currently, the Alberta government is taking an initiative to develop a sustainable resource and environmental management (SREM) policy in order to address issues related to multiple land use developments and its impact on the landscape and the environment[31]. One of the mandates of this project is to review upstream oil and gas development activities starting from exploratory drilling to reclamation and remediation. Projections of oil and gas drilling activities should be an important input to this initiative in terms of assessing the extent of exploration activities expected to take place on the landscape. Moreover, forecasts of drilling effort are useful in order to understand how economic policies will affect energy sector activities. For example, in the current debate related to the impact of royalty changes on oil and gas drilling activities, the Alberta Department of Energy estimated that a one percent increase in royalties would translate into about a reduction of 150 wells[32]. Forecasts given in the present study can be used to estimate the overall impact of current royalty changes on expected drilling activities. The spatial econometric model is also applied in the next chapter to analyze energy sector drilling activities in caribou habitat in anticipation of new regulations related to wildlife protection.

---

[31] Detailed explanation is given at Alberta's Commitment to Sustainable Resource and Environmental Management. www.srem.gov.ab.ca/pdf/1999_Commitment_document.pdf

[32] This finding is based on a technical report by the Alberta Department of Energy. Detailed information is found at: www.energy.gov.ab.ca/Oil/pdfs/RISConvTechRoyaltyImpact.pdf.

Chapter 6 Analysis of the energy sector's anticipation of environmental regulations

6.1 Introduction

Woodland caribou are one of the key indicators of a healthy and functioning boreal forest ecosystem. Historically woodland caribou in Alberta were listed as rare in 1984 and the provincial government placed the species on the "*Red List*" as a species at risk of extinction in 1991 (Dzus 2001). Recently, the Alberta government re-designated woodland caribou as a threatened species in 2001 (Dzus 2001). Caribou population growth is affected by a variety of factors such as predation, hunting, and industrial activities (Dzus 2001). Forestry and the energy sector are the two main industrial activities that resulted in the fragmentation and alteration of caribou habitat.

The main objective of this chapter is to test the hypothesis that the energy sector anticipated new environmental regulations in woodland habitat leading to increased exploration activities. A comparison between historical trends of exploration activities is made to analyze the intensity of exploration activities inside and outside caribou habitat. This comparison will help to test the hypothesis. In a similar setting, the term 'preemptive habitat destruction' is used in the literature to explain an action of pre-occupying land before a new regulation of conserving the land or forest area is implemented (Lueck and Michael 2003). Three statistical techniques are used to test this hypothesis. These techniques include multivariate regression, the difference in difference approach, and propensity score matching methods[33].

Using a multivariate regression method, Lueck and Michael (2003) assessed the popular notion that regulatory uncertainty induced by possible endangered species requirements influences landowners' decisions to cut timber quickly and foreclose potential endangered species habitat. Using data from 1984 to 1990 on over 1,000 individual forests in North Carolina, the authors show that owners of timberland close to land with colonies of protected red-cockaded woodpeckers (RCW) were more likely to harvest their timber when it is less mature. RCW rely on mature timber stands for nesting. By preventing the establishment of an old-growth pine stand, landowners can ensure that red-cockaded woodpeckers do not inhabit their land and avoid the endangered species act regulations that limit or prohibit timber harvest activity. In a similar study, Zhang (2004) found that land owners are 25 percent more likely to cut forests when they know or perceive that a RCW cluster is within a mile of their land. Moreover the study shows that possible regulatory intervention has a positive impact on a landowners' decision to employ clear-cutting instead of selective cutting as their harvesting method. The present study uses multivariate regression to test if the energy sector is drilling more wells in caribou habitat in anticipation of new conservation regulations. To date no empirical studies that we are aware of have been done on the behavior of the oil and gas sector and its anticipatory response to environmental regulation.

The second statistical technique used in this investigation is the difference in difference (DID) method. This method is defined as the difference in average outcome in the treatment group before and after treatment minus the difference in average outcome in the control group before and after treatment (Abadie 2005). The DID method of program evaluation is very common in labor market analysis (Card and Krueger 1994, Meyer et al 1995). For example, using this method Meyer et al (1995) examined the effect of workers' compensation on time out of work. Recent examples of the DID method in environmental economics include Hallstrom and Smith (2005) and Greenstone (2002, 2004)[34]. Hallstrom and Smith (2005) applied the DID approach to examine the response of housing values to information about new hurricanes. They used hurricane Andrew (the strongest hurricane to hit the U.S in 1992) as a case study. The study region was Lee County, Florida where this hurricane was a ‘near miss’. They hypothesized that Andrew conveyed risk information to homeowners in this county. The authors found that Andrew lead to a 19 percent decline in housing prices in ‘Special Flood Hazard Areas’ for Lee County, Florida. This finding implies that home buyers and sellers appear to have recognized the information conveyed by this storm and acted on it. The present study extends the literature on the application of the DID method using the case study of oil and gas exploration in caribou habitat. The findings from these analyses will help us to understand how the energy sector acts when there is an expectation of new regulations pertaining to caribou conservation.

---

[33] These methods are explained later in the chapter.

The third technique used to investigate the hypothesis is the propensity score matching method. The propensity score matching method has been applied in various types of policy analyses. For example, it has been used extensively to evaluate the impacts of employment programs in labor market analyses (Dehejia and Wahba 2002, Heckman et al 1997). Recent applications in environmental economics include Greenstone (2004) and List et al (2003) who evaluated the impact of the regulations specified by the US clean air act amendment on air pollution. Ferraro et al (2007) and Margolis et al (2004) have also applied this method to evaluate the effectiveness of the U.S. endangered species act on biodiversity conservation and to test the hypothesis of preemptive habitat destruction respectively. Margolis et al (2004) examined the extent of preemptive habitat destruction on more than 70,000 plots [illegible] potential Pygmy Owl critical habitat area in Pima County, Arizona. The study examined landowners' behavior in regards to applications for development permits when they expect that their land may be inside a critical habitat area. The authors tested the hypothesis that landowners expecting their land to be inside a critical habitat area would engage in more preemptive habitat destruction than those without such expectations. Their findings suggest that preemption is occurring at rates that are statistically and economically significant.

---

[34] The study by Meyer (1995) and a discussion paper by Greenstone and Gayer (2007) are good sources of the DID method and other program evaluation studies in a natural experiment setting.

Using the propensity score matching, Ferraro et al (2007) assessed the effectiveness of the U.S. endangered species act on native endangered terrestrial and freshwater vertebrates. They choose species that are present in one or more of the 50 states in the US. Their data consist of a sample of 135 listed species and 295 unlisted species. Their results show that listing a species under the endangered species act is, on average, detrimental to species recovery if not combined with substantial government funds. In the US, wildlife is a public resource while the habitat on which wild animals depend is often privately owned. The above findings show that private land owners are in most cases developing their land early before the endangered species act is implemented on the their land.

The present study differs from the previous studies in at least two ways. First, in the present study both the provincial land in Alberta and wildlife are mostly owned by the government. Hence, the findings from this study will help us to observe the action taken by the government in maintaining the balance between protection of caribou habitat and the exploration of oil and gas activities in the public land. Second, the present study is applying three different statistical methods to test the hypothesis instead of one method.

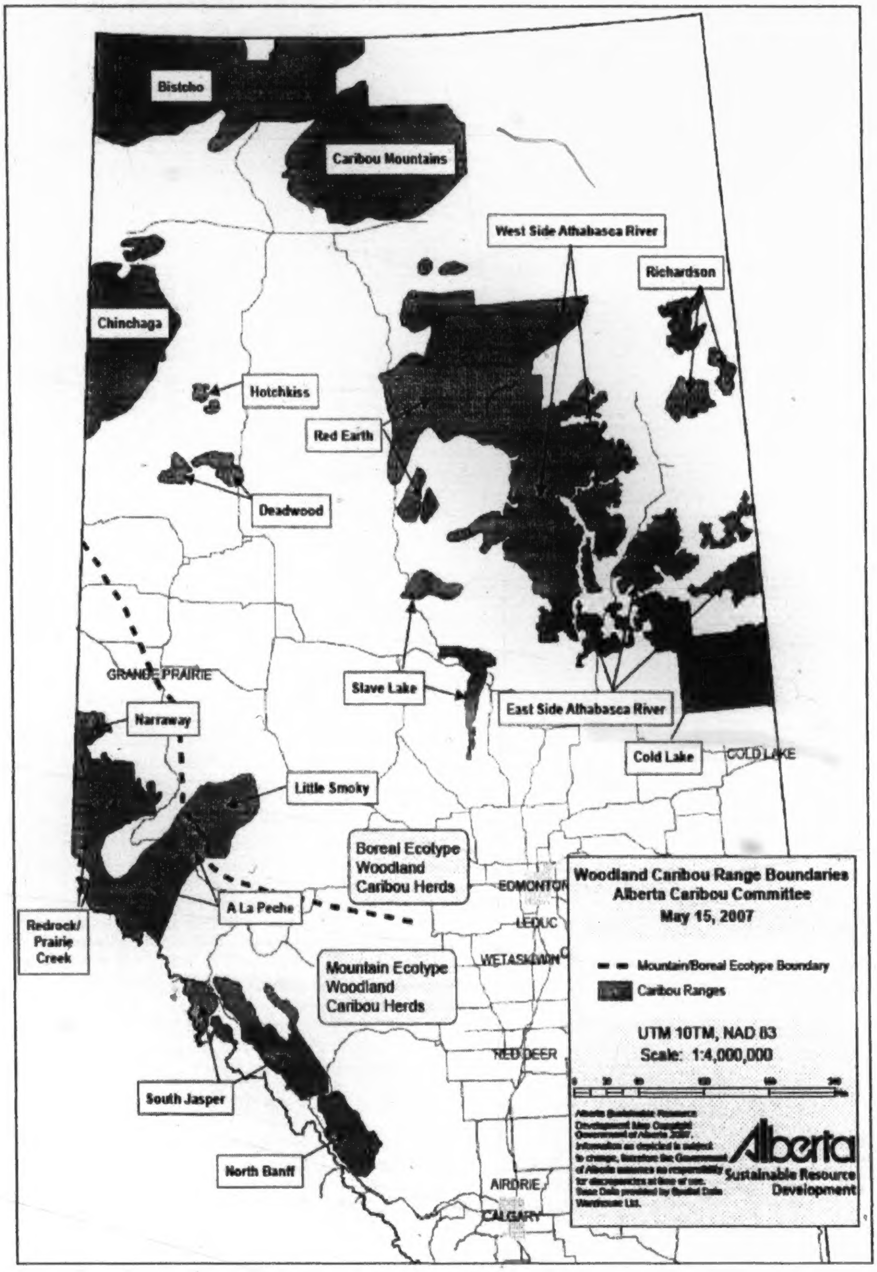
The rest of this chapter is organized as follows. Description of the hypothesis to be tested is explained in section two. In section three, caribou ranges in Alberta and a map of the study are described. Section four discusses data and descriptive analysis of historical exploration in woodland caribou habitat. In sections five to seven, the three statistical methods of testing the hypothesis are implemented. Section five deals with the multivariate regression analysis. This section examines drilling effort as a function of economic and geologic factors and includes caribou habitat as one of the explanatory variables. Section six discusses the difference in difference method (Meyer 1995). In this section exploration activities are compared between two time periods in caribou and non-caribou habitat. The third method, the propensity score matching, is discussed in section seven. Using propensity score matching (Rosenbaum and Rubin 1983); exploration rates for oil and gas are compared between caribou and non-caribou habitat provided that the two habitats are similar in many aspects such as geological, geographical, and ecological factors. A summary of the results and hypothesis tests are given in section eight and management implications of the findings of the study are discussed in the last section.

6.2 Description of the hypothesis test

The hypothesis is stated as follows: ‘the energy sector anticipates regulations related to caribou protection and increases exploration activities in caribou habitat prior to the implementation of the new regulations’. Historical events related to precautionary information regarding endangerment of caribou status are used to signal the anticipation of new regulations. Years 1991 and 2001 are chosen to represent historical events related to caribou population. The provincial government placed woodland caribou on the ‘Red List’ in 1991 and re-designated caribou as a threatened species in 2001 (Druz 2001). These historical events are used to test the hypothesis. The null hypothesis can be stated as average drilling in caribou habitat is the same before and after the event against the alternative that an increase in drilling effort is observed in caribou habitat after the event. Drilling efforts were observed before and after the historical events using the difference in difference method. This method allows us to compare drilling activities in caribou habitat versus non-caribou habitat before and after the specified historical event. The propensity score method is also used to compare drilling effort in caribou and non-caribou habitat assuming that the two sampled habitats are similar in their geological and geographical settings.

6.3 Caribou ranges and study area

Figure 6.1 Woodland caribou range boundaries in Alberta.


Bistcho
Caribou Mountains
West Side Athabasca River
Richardson
Chinchaga
Hotchkiss
Red Earth
Deadwood
GRANDE PRAIRIE
Slave Lake
East Side Athabasca River
Cold Lake
COLD LAKE
Narraway
Little Smoky
Boreal Ecotype Woodland Caribou Herds
EDMONTON
A La Peche
LEDUC
Redrock/ Prairie Creek
Mountain Ecotype Woodland Caribou Herds
WETASKIWIN
RED DEER
South Jasper
North Banff
AIRDRIE
CALGARY
Woodland Caribou Range Boundaries
Alberta Caribou Committee
May 15, 2007
Mountain/Boreal Ecotype Boundary
Caribou Ranges
UTM 10TM, NAD 83
Scale: 1:4,000,000
Alberta
Sustainable Resource Development

Source: Alberta Caribou Committee website

Figure 6.1 shows caribou range areas in Alberta. This map is taken from the official website of the recently formed Alberta caribou committee. This committee is a group of government, industrial, and academic partners in Alberta. They have worked for over a decade to integrate industrial activities in northern Alberta with the conservation of caribou habitat. There are two ecotypes of woodland caribou in Alberta. The first is the mountain ecotype caribou which are found in the foothills of the Rocky Mountains. In total there are five herds listed under this group. These include Narraway, Redrock_Praire Creek, A La Peche, South Jasper and North Jasper herds. The second is the boreal ecotype which is found in the northern and northeastern parts of Alberta. There are twelve herds listed under this group namely, Bistcho, Caribou Mountains, Chinchaga, Hotchkiss, Deadwood, Red Earth, Little Smoky, Salve Lake, West Side of Athabasca River (WSAR), East Side of Athabasca River (ESAR), Cold Lake, and Richardson (Alberta Caribou Committee website). Figure 6.2 shows histogram of herds and the approximate number of townships where the particular herd lives. The histogram shows that Caribou Mountains, Red Earth, the WSAR herds cover relatively large areas of habitat. For the purpose of this study all the caribou ranges, except those for the South and North Jasper herds, are included. The reason for excluding South and North Jasper herds is that these ranges are located along Jasper National Park.

Figure 6.3 shows the study area. This area covers all parts of the Northern and Foothills regions of Alberta north of township 50. Wood Buffalo National Park in northern Alberta and Jasper National Park along the Foothills region are excluded from the study. In total

3,566 townships are included in the study area. Out of these townships about 1167 townships are designated as habitat for woodland caribou.

Figure 6.2 Histogram of number of townships for each caribou herd

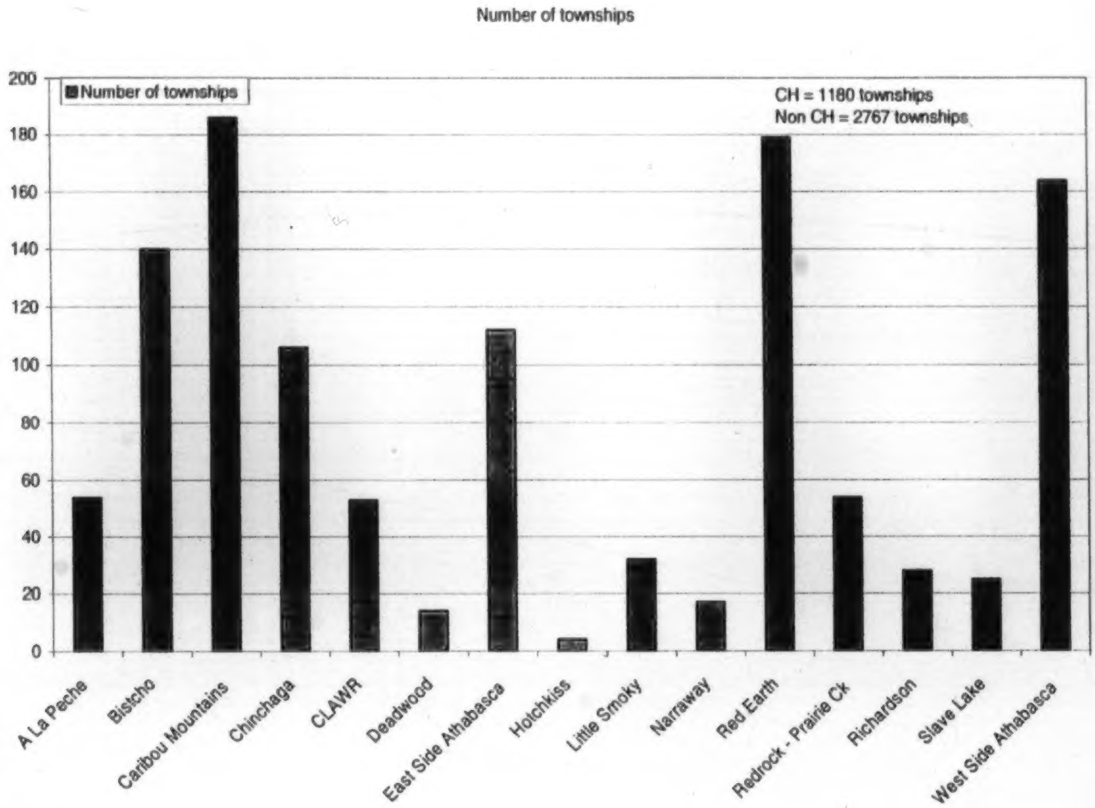
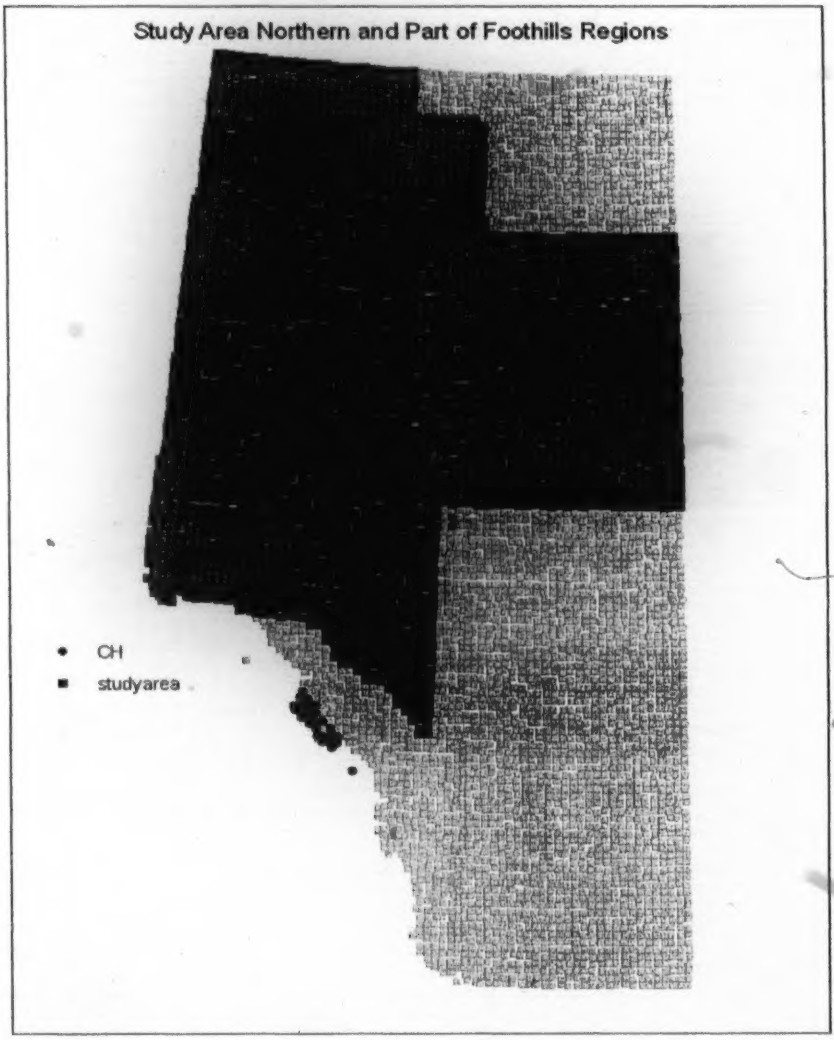


Figure 6.3 Map of study area


Study Area Northern and Part of Foothills Regions
● CH
■ studyarea

Note: CH refers to Caribou Habitat

6.4 Data and descriptive analysis of exploration in caribou habitat

6.4.1 Data

The data described in chapter three are used in this chapter. Some of the exceptions are that only a subset of the townships is included in this study (3,566 townships) and a dummy variable to represent caribou and non-caribou habitat is added to the variable list. In addition, spatial data on forest cover, water body, muskeg, roads, and the proportion of human disturbances are collected for the purpose of constructing a propensity score index[35]. The time period covers 1980 to 2004.

6.4.2 Descriptive analysis of exploration in caribou habitat

This section reviews average drilling density in caribou habitat using graphs and maps and compares the density of drilling in caribou versus non-caribou habitat. Figure 6.4 shows a trend of average drilling density over the study period. The upper line shows density of drilling in non-caribou areas and the bottom line shows drilling in caribou areas. For comparison purposes, density of drilling in caribou habitat after caribou herds along the Caribou Mountain and Bistcho area are excluded are shown in the middle line. The reason for excluding these herds is that historically less exploration activity was carried out in these areas and this may understate the intensity of drilling in other caribou habitats as a result of averaging. In other words these herds may be outliers. The graph shows that the trend of average drilling in caribou habitat is closer to the non-caribou habitat after these herds are excluded.

---

[35] Description of the data on forest cover and water-body and the propensity score matching method are discussed in section 6.7.

Figure 6.4 Trend of average well densities in caribou and non-caribou habitat[36]

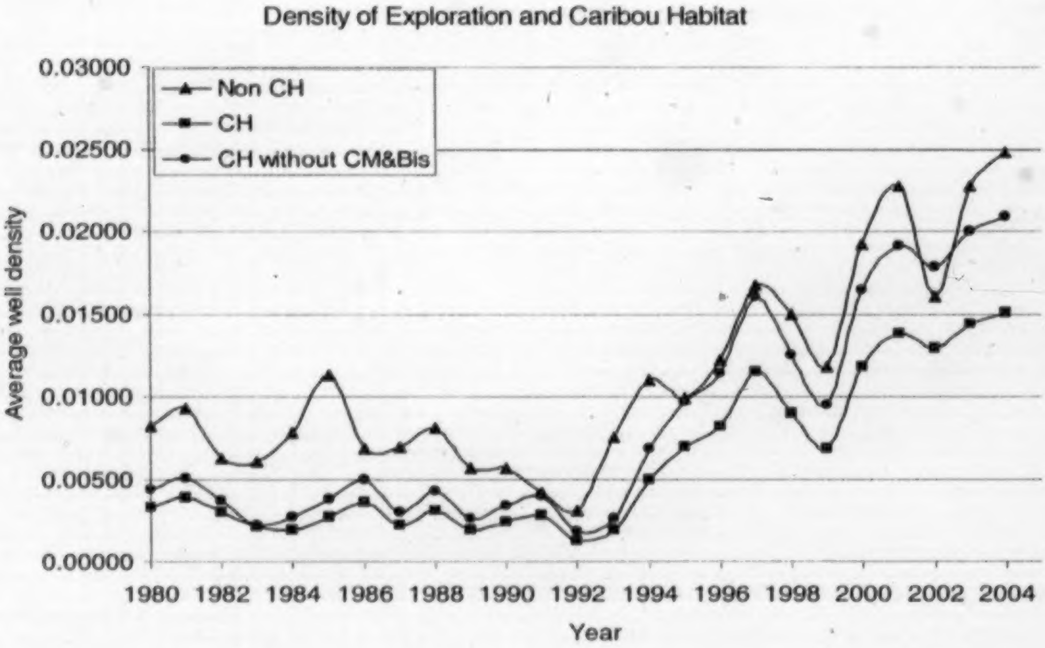


Figure 6.4 shows that density of average drilling is higher in non-caribou habitat throughout the study period. However, trends of average drilling for individual herds have mixed result. For example, Figure 6.5 shows the density of drilling in Little Smoky and Slave Lake herds and compares the trends with average drilling in non-caribou habitat. The graph shows that on average, the density of drilling in the Slave Lake area

was higher in most of the study period and drilling in the Little Smoky area was lower until 2001. Beyond 2001 it rises to the end of the time series. Another example is a comparison between average drilling in Cold Lake, West Side of Athabasca River (WSAR), East Side of Athabasca River (ESAR) and non-caribou habitat. This is shown in Figure 6.6.

---

[36] The units in the y-axis are average well density. Average *well density* is defined as the density of wells divided by number of townships in a given habitat per year. CH refers to caribou habitat and CM and Bis refer to caribou habitat ranges for Caribou Mountain and Bistcho herds.

Figure 6.5 Average well densities in Little Smoky, Salve Lake and non-caribou Habitat

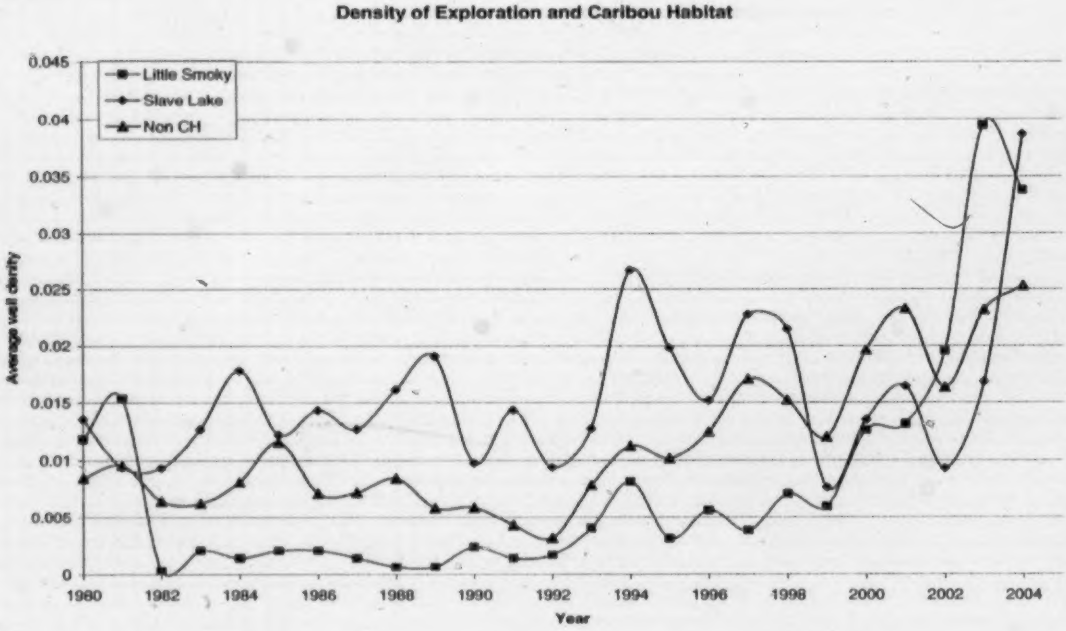
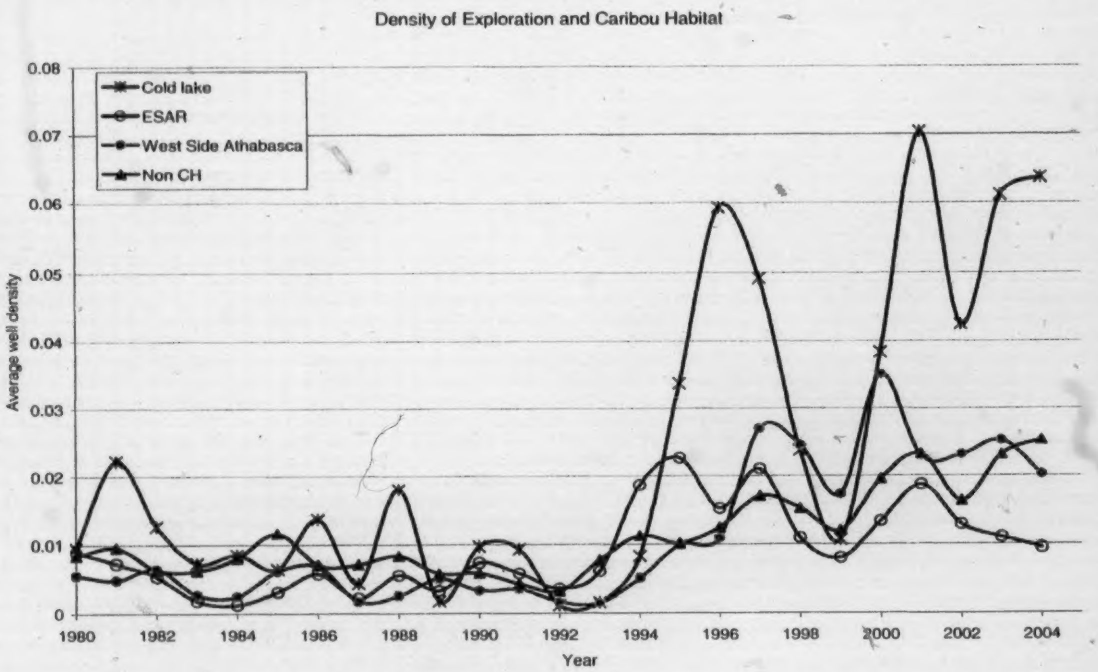


Figure 6.6 reveals that drilling activities in the Cold Lake area were higher than the non-caribou area throughout most of the study period. Specifically, the density of drilling has increased substantially starting in 1995. For the case of WSAR and ESAR, average

densities of drilling were lower than the non-caribou areas until 1994. After 1994 average drilling increased especially in the WSAR[37].

Figure 6.6 Average well densities in Cold Lake, WSAR, ESAR, and non-caribou Habitat



An alternative way of comparing drilling activities in caribou ranges with non-caribou habitat is to group the herds on a regional basis. Caribou herds can be categorized into four different regions based on their proximity to each other. For the purpose of this study these regions are named North East, North West, Far North West and West Central regions[38]. The North East region covers six herds. These are ESAR, WSAR, Slave Lake, Red Earth, Cold Lake and Richardson. The Far North West region includes Bistcho and Caribou Mountain herds. The North West region includes Chinchaga, Hotchkiss, and

---

[37] Graphical comparison of drilling activities for other individual herds is attached as Appendix 6.1

[38] This classification is similar to the PSAC (Petroleum Services Association of Canada) areas and the Alberta caribou recovery plan team (2005).

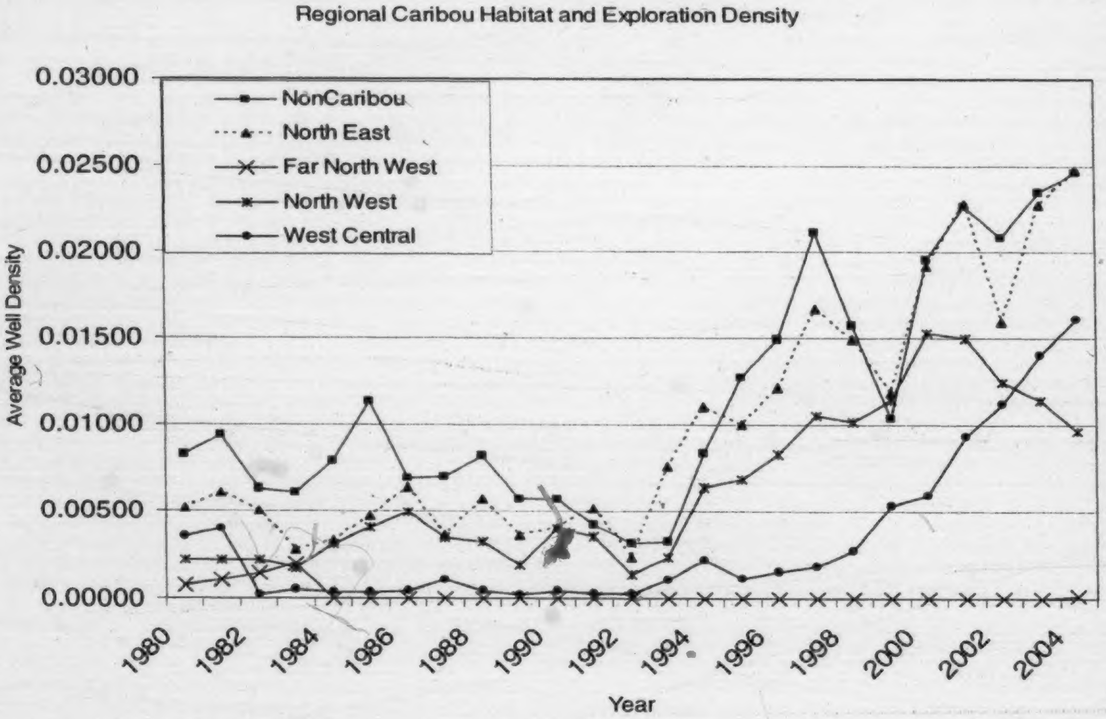
Deadwood herds. The fourth one is the West Central region that includes four herds – Little Smoky, Redrock_Prairie Creek, A La Peche and Narraway. Figure 6.7 shows a comparison of average drilling density between these regions and non-caribou areas. This figure indicates that average drilling density was higher in the non-caribou areas until 1992. The trend of drilling after the year 1992 shows that average drilling in the North East regions rises and in some years it is either higher or equal to the non-caribou areas. In the West Central and North West regions, average density of drilling is lower than the non-caribou areas but the trend after the year 1995 indicates that drilling is increasing through time especially in the West central region.

The above analyses show that comparing average drilling in caribou and non-caribou areas can give misleading results if all the herds are treated as one group. Aggregating all the caribou herds as one group and comparing the results with areas of non-caribou habitat understates the density of drilling due to the heterogeneous nature of the herds. Density of average drilling varies not only among the individual herds but also between different time periods within each herd. For example, little drilling activity was carried out in the Caribou Mountains area and intensive exploration was carried out in the Cold Lake area. Furthermore, the density of drilling in the Richardson herd was low before the year 2000 but a tremendous increase was observed after 2001[39]. Hence, the above descriptive analyses are not adequate to compare exploration activities between the two habitats. A more advanced statistical analysis that takes in to account other factors of exploration and the spatial and temporal trends of drilling are discussed in the next section.

---

[39] See the graphs in Appendix 6.1 that shows trends of drilling for the Richardson herd.

Figure 6.7 Average density of drilling aggregated into four regions



6.5 Multivariate regression

6.5.1 Multivariate model

The model developed in chapter 2 is used to compare exploration for oil and gas in caribou and non-caribou habitat. A dummy variable for caribou habitat (CH) is included on the right hand side of the equation. This variable that takes a value of one if a given township falls in caribou habitat and zero otherwise. For convenience the model is re-written below:

$$w_{it} = F\left(S_{it-1}, A_{it-1}, A^2{}_{it-1}, P_t, T_t, U_t, CH_{it} / \beta_k, \varepsilon_{it}\right) \qquad 6.1$$

All the variables in equation 6.1 are as defined in chapter 2. The spatial lag version of the model is used to estimate this equation. The sign of the coefficient for the CH variable is an empirical question. A positive sign would indicate that density of exploration is higher in caribou habitat than in non-caribou habitat after controlling the other explanatory variables. The next sub-section discusses results of the regression.

6.5.2 Results

Three different versions of the multivariate regression model are estimated. The first model is a simple multivariate regression equation where caribou habitat is included on the right hand side of the equation. The second model is multivariate regression model where a dummy variable is created for each herd and non-caribou habitat is used as a baseline. The third model is a regression equation where caribou habitat is included in the equation based on PSAC (Petroleum Services Association of Canada) areas[40]. According to this classification Alberta is divided in to seven sub-regions. The PSAC area map was originally designed based on geological similarities and the kind of drilling activity it generated and is used for the purpose of estimating costs of drilling. Out of the seven sub-regions, caribou habitats are found in four regions (Regions 1, 2, 6, and 7). These regions are used in the multivariate regression model.

---

[40] Refer Appendix 6.2 for the map of PSAC areas.

Table 6.1 Results for the multivariate regression model of caribou habitat: Model 1

Model 1: Simple multivarite regression

Variable	Coef	(t-stat)
Constant	-0.010	(-14.17)
Cumulative Wells	0.054	(74.66)
Cum. Wells Squared	-0.003	(-15.50)
Lag. Success Rate	0.018	(36.80)
Lag Price-1 Period	0.003	(11.82)
Capacity Utilization	0.017	(14.50)
3D Seismic Dummy	-0.002	(-3.32)
Time Trend	0.000	(-0.03)
Caribou Habitat	0.001	(4.08)
W Dep. Variable*	0.045	(8.99)
N	98375	
R-squared	0.19	
Log-L.hood	183938.3	

Note: t- statistics in parenthesis

Table 6.1 provides results for the first version of multivariate regression model. The signs and the statistical significance of most of the coefficients are similar to the spatial lag models estimated in chapter 4. Hence, there is no need to repeat their interpretations here. The main parameter of interest in this model is the coefficient for caribou habitat. This coefficient is positive and significant. This suggests that on average more drilling activities are carried out in caribou habitat than in non-caribou habitat after controlling for the other explanatory variables.

The second version of the multivariate regression model is estimated using a dummy variable for each of the 15 caribou herds. In total there are 16 variables that represent habitat in which 15 dummy variables are created for each herd and non-caribou habitat is used as a baseline variable. The regression model also includes time period dummy variables where the period from 1990 to 1992 is selected as a baseline period. Four time dummies are created to capture temporal variations in drilling. An interaction term between each caribou herd and the four time period dummies are then created to observe drilling activities in each caribou herd over the specified time period. These interaction terms are our main variables of interest. A summary of the results is given in Table 6.2 and detailed results are attached as Appendix 6.3. The results show that the Cold Lake and WSAR herds have positive and significant coefficients. This shows that on average drilling activities in these herds are higher than in non-caribou areas. The results for the Richardson herd, north east part of Alberta, show that the coefficient for the recent period is positive and significant. Oil and gas exploration in Caribou Mountains and Bistcho herds are always lower compared with non-caribou habitat. The results for most of the other herds are either insignificant or mixed results where in some time periods positive and statistically significant coefficients are observed and in other time periods the coefficients are not significant.

Table 6.2 Summary of results for the multivariate regression model estimated using dummy variables for each herd: Model 2 herd by herd model.

Herd name	Significance of the coefficients
Cold Lake	Positive and significant
WSAR	Positive and significant
Richardson	Mixed results but significant
Caribou Mountain	Negative and significant
Bistcho	Negative and significant
ESAR	Negative and significant
Narraway	Positive but insignificant
Red Earth	Negative but insignificant
Hotchkiss	Negative but insignificant
Deadwood	Negative but insignificant
A La Peche	Negative but insignificant
Redrock/P.Creek	Negative but insignificant
Little smoky	Mixed results but insignificant
Chinchaga	Mixed results but insignificant
Slave Lake	Mixed results but insignificant

Note: WSAR and EASR refer to west side and east side of Athabasca River. Detailed results are found in Appendix 6.3

The third version of the multivariate regression model is estimated using the classification by PSAC areas. A map of the area is attached as Appendix 6.2. Four equations are estimated for regions 1, 2, 6, and 7. For each region two different models are estimated. The first model is a simple regression model where caribou habitat is included as a dummy variable. The second model is a regression model where an interaction between caribou habitat and a time dummy is included in the equation. In the second case the year from 1980 to 1982 is used as a baseline year and six time dummy variables are created on a three year time intervals. In these models our main variables of interest are the interaction terms between caribou habitat and the time dummies. These variables indicate the density of drilling in the specified caribou region over the given time interval. Results for each region are given in Table 6.3.

For model 1, the results show that significantly higher drilling is observed in caribou habitat in regions 1 and 6. In region 7 drilling in caribou habitat is less than in non-caribou habitat. Region 7 includes herds such as Caribou Mountain and Bistcho. These results are consistent with the previous herd by herd model. For region 2 the coefficient for caribou habitat is positive but not statistically significant. The results from model 2 show that significant drilling is observed in region 6 almost throughout the study period. In region 1 drilling has increased in recent years compared with the base year. For region 7, even though the coefficient is not statistically significant, drilling in caribou ranges is less than in non-caribou ranges in the years between 2001 and 2004. Overall the results in Table 6.3 are consistent with the results in Table 6.2. This consistency is attributed to the fact that region 6 includes Cold Lake and WSAR herds which have positive and significant coefficients and region 7 includes Caribou Mountain and Bistcho herds which have negative and significant coefficients.

Table 6.3 Multivariate regression results for each region: Model 3 - PSAC regional model

Variable	Region 1		Region 2		Region 6		Region 7	
	Model 1 Coef.	Model 2 Coef.	Model 1 Coef.	Model 2 Coef.	Model 1 Coef.	Model 2 Coef.	Model 1 Coef.	Model 2 Coef.
Constant	-0.003*	0.000	-0.023*	-0.018*	-0.019*	-0.012*	-0.002*	-0.005*
Cumulative Wells	0.093*	0.094*	0.039*	0.038*	0.080*	0.079*	0.035*	0.036*
Cum. Wells Squared	0.122*	0.116*	-0.002	-0.002	-0.007*	-0.007*	0.002*	0.001*
Time Trend	0.000	0.000	0.000	0.000	-0.001	0.001	-0.000*	0.001
Lag. Success Rate	0.001*	0.010*	0.023*	0.023*	0.012*	0.011*	0.025*	0.025*
Lag Price-1 Period	0.001*	0.000	0.010*	0.006*	0.005*	0.003	0.001*	0.001*
Capacity Utilization	0.001	0.000	0.031*	0.031*	0.027*	0.023*	0.008*	0.009*
Caribou Habitat	0.001*	0.000	-0.001	-0.001	0.004*	-0.001	-0.001*	-0.001
Time Dummy 83_85		0.000		0.000		-0.008*		0.002*
Time Dummy 86_88		0.000		0.001		-0.008*		0.000
Time Dummy 89_91		0.000		-0.002		-0.007		-0.002
Time Dummy 92_94		0.000		-0.003		-0.013*		-0.002
Time Dummy 95_97		0.001		-0.001		-0.008		-0.002
Time Dummy 98_00		0.002		0.000		-0.010		-0.002
Time Dummy 01_04		0.000		0.007		-0.006		-0.002
*CH*Time 83_85*		0.000		-0.003		0.004		-0.002*
*CH*Time 86_88*		0.000		0.000		0.006		0.000
*CH*Time 89_91*		0.000		0.005		0.005		0.001
*CH*Time 92_94*		0.000		0.003		0.006		0.000
*CH*Time 95_97*		-0.001		-0.003		0.009		0.001
*CH*Time 98_00*		0.000		-0.001		0.008		0.000
*CH*Time 01_04*		0.003*		-0.003		0.006		-0.001
*W*Dep. Variable*	0.034	0.027	0.056*	0.054*	0.042*	0.040	0.040*	0.040
N	3575	3575	12975	12975	24500	24500	50050	50050
R-Squared	0.22	0.23	0.26	0.26	0.17	0.17	0.23	0.23
Adj. R-Squared	0.22	0.23	0.25	0.26	0.16	0.17	0.23	0.23
Log-Likelihood	12802.6	12815.5	25378.0	2538890	33833.2	33845.6	119298.7	119304.6

Note: * shows statistically significant coefficients at 5% level of significance.

Even though the results obtained from the multivariate regression models are informative, there are a number of limitations to this approach. For instance, the positive coefficient for caribou habitat does not specifically indicate the time period where significant drilling activities start to dominate in this habitat. This specification provides only an average indication of drilling activities through out the study period or on the specified time period set as a time dummy variable. To overcome this limitation, the difference in difference method is employed. The advantage of the difference in difference method is that a specific intervention period can be identified so that drilling activities are compared in caribou versus non-caribou habitat before and after the intervention period. Based on the multivariate regression models, it may not be appropriate to conclude at this point that the energy sector is engaged in "excessive" drilling in caribou areas anticipating new regulations.

The other limitation of the multivariate regression method is that it does not specifically or rigorously control for differences in caribou and non-caribou habitat in terms of their geological and ecological characteristics. Even though a number of explanatory variables are included in the model, these variables are useful only to control for the factors that affect exploration activities. The propensity score matching method is a better approach to use since it creates a probability score where one can match two groups with similar characteristics based on a number of geological or ecological criteria. A comparison can then be made on the impact of new information or regulation both through time and space. The next two sections discuss the difference in difference and the propensity score matching methods and the results based on these methods.

6.6 The difference in difference method

6.6.1 Introduction

The difference in difference (DID) method is one of the popular tools of applied research in labor economics. This method is often used to evaluate the effects of public policy interventions (Abadie 2005). The main components of the DID estimator are the treatment group, control group, the treatment or public intervention, and the treatment effect or outcome. In the present study the treatment group is caribou habitat, the control group is non-caribou habitat, the treatments or public interventions are historical events related to caribou conservation planning or public policy related to oil and gas exploration on caribou habitat. The treatment effect is the average density of oil and gas exploration after the intervention. The DID estimator can be defined as the difference in average exploration activities (well density) in caribou habitat before and after a historical event (e.g. caribou listed as 'endangered') minus the difference in average exploration in non-caribou habitat before and after the historical event. Hence it is named the 'difference in difference' method.

Two historical events (scenarios) are used to examine the response of the energy sector to exploration activities in caribou habitat compared to non-caribou habitat. The first event is the period when the provincial government placed woodland caribou on the 'Red List' as a species at risk of extinction in 1991. This period is used as a benchmark to examine the response of energy sector exploration activities in caribou habitat after caribou was listed as a species at risk. The second event is the year 2001 when the provincial government re-designated woodland caribou as a threatened species. This analysis will help us to examine how the energy-sector was responding in regards to exploration activities after the two historical events were announced. For each scenario, exploration activities before and after the event are examined. Mathematical formulation of the model and the method of analysis are explained in the next sub-section.

6.6.2 Model and methods

The basic logic behind the DID method is to model the treatment effect by estimating the outcome measures between two time periods for both the treated observations and the control groups. Specifically, the effect of the historical event (e.g. year 2001, caribou listed as threatened species) is estimated in terms of density of drilling in caribou habitat compared with drilling in non-caribou habitat. The underlying model of the outcome variable can be written as[41]:

$$w_{it}^{j} = \alpha + \alpha_1 TD + \alpha_2 CH^{j} + \alpha_3 (TD * CH^{j}) + \varepsilon_{it}^{j} \tag{6.2}$$

where w_{it}^{j} is the outcome variable (drilling density) at time t in township i and the index j represents a group, j = 1 for caribou habitat and 0 for non-caribou. TD refers to a time-dummy where it takes a value of 1 after a certain event and 0 before the event. CH is a dummy variable for caribou habitat where it takes a value of one for caribou habitat, 0 otherwise. TD*CH is the interaction between the time dummy and habitat dummy variables. The coefficients α_1 and α_2 capture time specific and time-invariant differences between the groups respectively. α_3, the parameter of interest, shows the effect of the treatment (an event) on exploration activities in caribou habitat. A positive coefficient would indicate that more exploration activities are carried out after the warning sign of caribou's endangerment status or after the release of information related to caribou habitat conservation. The final term, ε_{it}^{j} is the error or disturbance term for each unit at each time period with the assumption of zero mean and constant variance.

---

[41] The formulation of the DID model is based on Meyer (1995).

Assuming that the model is correctly specified, the error term is on average zero, $E(\varepsilon_{it}^{j}) = 0$, and it is uncorrelated with the other variables in the equation; the expected values of the average outcomes in equation 6.2 can be written as:

$$E(w_{TD=0}^{1}) = \alpha + \alpha_2$$

$$E(w_{TD=1}^{1}) = \alpha + \alpha_1 + \alpha_2 + \alpha_3$$

$$E(w_{TD=0}^{0}) = \alpha$$

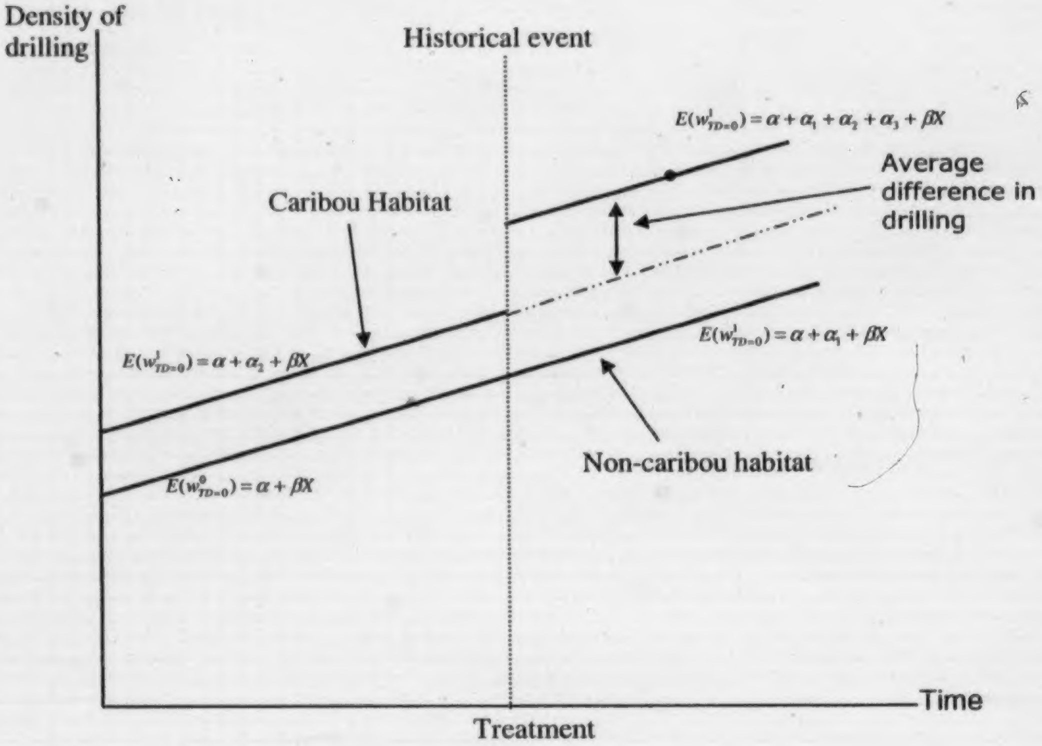
$$E(w_{TD=1}^{0}) = \alpha + \alpha_1$$

where E shows the expected or average values over townships, subscripts denote the time period, and superscripts denote the group. Equation 6.2 is called the DID method because the unbiased estimate of α_3 can be obtained by using the following expression:

$$\hat{\alpha}_3 = E(w_{TD=1}^{1} - w_{TD=0}^{1}) - E(w_{TD=1}^{0} - w_{TD=0}^{0}) \qquad 6.3$$

Equation 6.3 shows the average treatment effect. A graphical explanation of the above discussion is given in Figure 6.8.

Figure 6.8 Graphical explanation of the difference in difference (DID) method.


Density of drilling
Historical event
$E(w^1_{TD=0}) = \alpha + \alpha_1 + \alpha_2 + \alpha_3 + \beta X$
Average difference in drilling
Caribou Habitat
$E(w^1_{TD=0}) = \alpha + \alpha_2 + \beta X$
$E(w^1_{TD=0}) = \alpha + \alpha_1 + \beta X$
Non-caribou habitat
$E(w^0_{TD=0}) = \alpha + \beta X$
Time
Treatment

The DID model is based on the assumption that the average change in the outcome is presumed to be the same for both groups if the treated group (caribou habitat) was not designated as caribou area. The other issue that needs to be addressed is the incorporation of other explanatory variables that affect drilling activities in both caribou and non-caribou habitat. The simple model given in equation 6.2 is not sufficient to capture the dynamics that occur in the real world. The traditional way to accommodate covariates in the DID model is to introduce them linearly in equation 6.2 (Meyer 1995). Technically speaking the DID model is a combination of the model developed in the previous section (equation 6.1) and equation 6.2. This equation can be written as:

$$w_{it}^{j} = \alpha + \alpha_1 TD + \alpha_2 CH^{j} + \alpha_3 (TD * CH^{j}) + \beta X_{it} + \varepsilon_{it}^{j} \qquad 6.4$$

where X indicates a vector of the explanatory variables discussed in equation 6.1 and β is a vector of coefficients for X. Equation 6.4 is estimated using the spatial lag model developed in chapter 2 in order to account for the spatial behavior of oil and gas drilling activities in each township. Meyer (1995) comments that the inclusion of the control variables helps to adjust for observable differences between the observations in the two groups and it may also improve the efficiency of the estimate of α_3 by reducing the residual variance. This is based on the assumption that β is equal across the caribou and non-caribou habitat groups. Coefficient estimates for equation 6.4 and discussion of the results are given in the next sub-section.

6.6.3 Results for the difference in difference model

Results of the DID model for the two cases or scenarios are given in Table 6.4 below. The sign and significance of the coefficients for the explanatory variables are consistent with the results obtained in chapter four and with the results of the multivariate regression model discussed above. In the present case the main parameter of interest is the coefficient for the interaction between caribou habitat and the time dummy. A positive coefficient for this variable explains that, after controlling for the factors that affect oil and gas exploration, the average density of drilling is higher in caribou habitat after the historical event. The result for the first scenario shows that the average density of drilling is higher in caribou habitat after the provincial government placed woodland caribou on the 'Red List' as a species at risk of extinction in 1991. This coefficient is statistically significant. The result for the second scenario indicates that the coefficient of caribou habitat after caribou was re-designated as a threatened species in 2001 is positive and significant. Based on these results we can conclude that the energy sector is not slowing down its exploration activities in caribou habitat after a cautionary statement about the status of caribou population is announced.

The DID method has at least two limitations. The first one is similar to the limitation discussed for the case of multivariate regression. The DID method does not rigorously control for differences in caribou habitat and non-caribou habitat in terms of their geological and ecological characteristics. The second limitation is that the DID method can suffer from functional form misspecifications like other parametric approaches. The propensity score matching method is used to address these issues. This method is a quasi-parametric approach since propensity scores are estimated parametrically while the treatment effects are non-parametrically determined.

Table 6.4 Results based on the difference in difference model for the two scenarios:

	Scenario 1*	Scenario 2*
Variable	Coef	Coef
Constant	-0.014	-0.015
	(-14.98)	(-13.08)
Cumulative Wells	0.054	0.054
	(73.35)	(72.78)
Cum. Wells Squared	-0.003	-0.003
	(-14.82)	(-14.61)
Lag. Success Rate	0.017	0.017
	(34.03)	(34.04)
Lag Price-1 Period	0.003	0.005
	(10.69)	(9.43)
Capacity Utilization	0.015	0.012
	(12.22)	(10.70)
Time Trend	0.003	0.001
	(8.58)	(5.53)
Caribou Habitat (CH)	0.000	-0.003
	(0.39)	(-4.21)
Time Dummy	-0.004	0.001
	(-7.89)	(4.04)
*CH*Time Dummy* (α_3)	0.002	0.000
	(3.25)	(3.05)
*W*Dep. Variable*	0.038	0.042
	(8.22)	(8.65)
N	98375	98375
Adj. R-squared	0.18	0.18
Log Likelihood	183797.7	183840.8

Note: t –statistics are in parenthesis. * Scenario 1: Historical event when caribou was placed on the 'Red List' in 1991 and Scenario 2 when caribou was re-designated as threatened species in 2001

6.7 Propensity score matching

6.7.1 Introduction

To this point the comparison between exploration activities in caribou and non-caribou habitat has been implemented using multivariate regression and the difference in difference methods. Having identified the shortcomings of these methods, an alternative is to use a semi-parametric matching procedure: propensity score matching. Propensity score matching refers to a class of multivariate methods used in comparative studies to construct treated and matched control samples that have similar distributions on many covariates (Rosenbaum and Rubin 1985). In the context of the present study the basic logic behind this method is to choose townships from non-caribou habitat that have similar characteristics, in terms of geological and ecological factors, to townships in caribou habitat. A comparison between these habitats is then made to examine the extent of exploration of oil and gas in caribou habitat in a given time period. Theoretical and methodological formulations of this method are discussed in the following section.

6.7.2 Methodology[42] and data

The main outcome variable of interest is the difference in oil and gas exploration density between caribou and non-caribou habitat. Let the treatment condition be denoted by CH = 1 for caribou habitat townships and CH = 0 for non-caribou habitat townships. In the program evaluation literature caribou habitat townships are called the treated group and non-caribou habitat townships are called the control group. Let the impact variable of interest be denoted by w^1 for exploration of oil and gas activities in caribou habitat and w^0 for exploration activities in non-caribou habitat. The objective is to then to estimate the difference in exploration activities in caribou and non-caribou habitat. If the treatment is assigned randomly (CH and non-CH), then it can be assumed that the covariates and un-observables do not differ in any systematic way between the two groups. In other words, they come from the same distribution. In this case, to estimate the average difference in exploration activities one can compare the outcome level after an intervention or after a regulation is set out. Hence, the average treatment impact in a randomized setting can be calculated as:

---

[42] Theoretical background and practical guidelines of implementing propensity score matching is explained in detail in Vinha (2006) and Caliendo and Kopeinig (2005).

$$E(w) = E(w^1 \mid CH = 1) - E(w^0 \mid CH = 0) \qquad 6.5$$

where E(w) is the mean difference in exploration between caribou and non-caribou habitat based on the assumption that townships in caribou habitat would have had, on average, the same outcome level as the non-caribou group had they been assigned to these group, i.e. $E(w^0 \mid CH = 1) = E(w^0 \mid CH = 0$.

In the case of Alberta's caribou habitat the above assumption does not hold because habitats are not assigned randomly. Therefore, drilling activity in non-caribou habitat is not a valid counterfactual for drilling activity in caribou habitat. The main challenge is then to construct a control group (non-caribou habitat) that is similar in covariates to the treatment group (caribou habitat). These covariates should capture variables that reflect assignment to the treatment and control groups and those that influence the outcome measure. The average treatment effect is then calculated as the difference in the average outcome of drilling activities in caribou habitat and the 'matched' non-caribou habitat with a similar set of covariates.

Matching on the covariates guarantees that the two groups have similar distributions of covariates and a treatment impact that mimics that of a randomized experiment (Vinha 2006). When the number of covariates is large the method of matching is not feasible since it becomes extremely hard to capture all the covariates simultaneously. As a solution to this Rosenbaum and Rubin (1983) show that it is not necessary to match groups based on the vector of covariates per se; matching on balancing scores, such as the propensity score (PS) is sufficient. The propensity score is defined as the probability of being assigned to the treatment group given the covariates. This probability is an index of all the covariates and effectively compresses the multi-dimensional vector of covariates in to a simple scalar. Based on this index equation 6.5 can then be written as:

$$E(w) = E(w^1 \mid PS, CH = 1) - E(w^0 \mid PS, CH = 0) \qquad 6.6$$

Equation 6.6 explains that the average treatment effect (exploration of oil and gas after a certain event) is the difference between the average drilling in caribou habitat minus the average drilling in a matched non-caribou habitat after conditioning on the propensity score (PS).

There are a number of practical issues that need to be addressed when implementing matching using propensity score methods. The first issue is how to estimate the propensity score index and what variables to include as covariates. The second issue is to determine which matching algorithm to use. Another issue that needs to be addressed is to make sure that the resulting control group sample (non-caribou habitat) is similar in the observable covariates to the treated group (caribou habitat) after matching is performed. That is whether or not the two samples are balanced after the appropriate matching algorithm has been applied to obtain the counterfactuals for each treatment observation. These issues and how they are addressed in the present study are discussed in the next paragraphs.

In practice the first step is to estimate a propensity score index using a binary choice model (logit or probit) where the dependent variable is whether or not the observation is in the treatment group (caribou habitat = 1 or non-caribou habitat = 0). According to Dehejia and Wahba (2002) the role of the propensity score is only to reduce the dimensions of the conditioning and as such it has no behavioral assumptions attached to it. In this study a logit model is used to estimate the propensity score index as follows[43]:

$$\Pr(CH = 1) \mid X) = \frac{e^{\beta h(X)}}{1 + e^{\beta h(X)}} \qquad 6.7$$

where CH is caribou habitat and h(X) is made up of linear and higher order terms of covariates. Ecological variables such as the percentage of forest cover, water bodies, muskeg, roads, and other human disturbances are used for each township. Proxies for geological variables such as cumulative wells, cumulative wells squared, and average success rate of finding oil or gas are used to capture the extent of available reserves, depletion of reserves, and resource discovery rate. One of the limitations of the propensity score matching application is lack of temporal and spatial data on reserves at a township level. Reserves of oil and gas are a major factor that drives exploration activities. Hence, cumulative number of wells is used as a proxy variable for reserves. Cumulative wells are also a good proxy for energy sector infrastructure such as pipelines and other facilities. Geographical variables such as latitude and longitude are also included in the logit model. The inclusion of these variables is to ensure that the two groups of townships are similar in many aspects.

---

[43] Detailed procedures of estimating the PS are given in the appendix on Dehejia and Wahba (2002).

The second practical issue is which matching algorithm to use. For the purpose of this study a one to one matching method is used[44]. This algorithm is commonly called nearest neighbor matching where the logical match for each treatment observation is the control observation with the closest propensity score. The advantage of using this algorithm is that it reduces the bias that is introduced when the matched pairs are less similar in their probability of receiving treatment (Vinha 2006). An important feature of the nearest neighbor matching is that, after the units are matched, the unmatched comparison units are discarded and are not used in estimating the treatment impact. In this case the treatment impact is given by the following equation:

$$E(w) = \frac{1}{N}\sum_{i=1}^{N}(w_i^1 - w_j^0) \qquad 6.8$$

---

[44] Other options are Kernel or Caliper matching methods. Kernel matching refers to matching estimators that use weighted averages of observations in the control group and Caliper matching refers to imposing a tolerance level on the maximum propensity score distance.

where N is the number of observations in each treatment and control group, w_i^1 is exploration activities in caribou habitat in township i and w_j^0 is exploration activities in non-caribou habitat in township j which has the closest propensity score to observation i. The third practical issue is to make sure that the resulting matched control groups are similar in the covariates to the treated group. In other words to counter-check that the explanatory variables used to estimate the propensity score are properly matched. This is implemented by checking the statistical significance of the mean difference of the covariates for the two groups using t-tests. According to Dehejia and Wahba (2002) if the two samples are not similar then additional higher order terms such as squares of the covariates or interaction terms of the covariates need to be included in the construction of propensity scores until the two samples are reasonably matched. In the next section results for the propensity score matching are presented using graphs, maps and tables.

6.7.3 Propensity Score Matching Results – Part I

The first step in the process of propensity score matching is to estimate the propensity score index using a logit model. A cross-sectional data set was built to estimate equation 6.7. This data set is composed of cumulative wells, cumulative wells squared, and average *success rate* for the period of 1980 - 2001. A total of 3566 townships are included in this data set. The year 2001 is chosen as a bench mark to estimate the impact of drilling activity in caribou habitat after the provincial government re-designated woodland caribou as a threatened species. Based on this estimate the response of the energy sector to this historical event is analyzed for the subsequent four years (2001 to 2004). For this data set time-invariant variables, at least for the given study period, are included. These variables include the percentage of forest cover for each township, proportion of water bodies, muskeg, roads, and human disturbance such as buildings for each township. In addition, geographical variables that represent the latitude and longitude location of each township are included in the model. Results for the logit model are given in Table 6.5 below.

Table 6.5 Results of the logit model used constructing the propensity score index

Variable	Coef	t-stat
Dependent variable : CH		
Cumulative Wells	-0.550	(-1.95)
Cum. Wells Squared	0.060	(1.26)
Avg. Success Rate	0.600	(4.99)
Forest Cover (%)	-17.10	(-6.19)
Forest Cover Squared	8.850	(10.93)
Water Body Cover (%)	-11.08	(-3.71)
Muskeg Cover (%)	-4.430	(-1.54)
Human Disturbance (%)	-10.14	(-3.54)
Roads	-0.060	(-9.45)
Roads Squared	0.000	(2.83)
Area of Township	- 0.002	(-0.55)
Easting (Latitude)	-1.870	(4.05)
Northing (Longitude)	13.78	(3.15)
N	3566	
R-squared	0.28	
Log-Likelihood	-1679.8	

Note: t-statistics in parenthesis

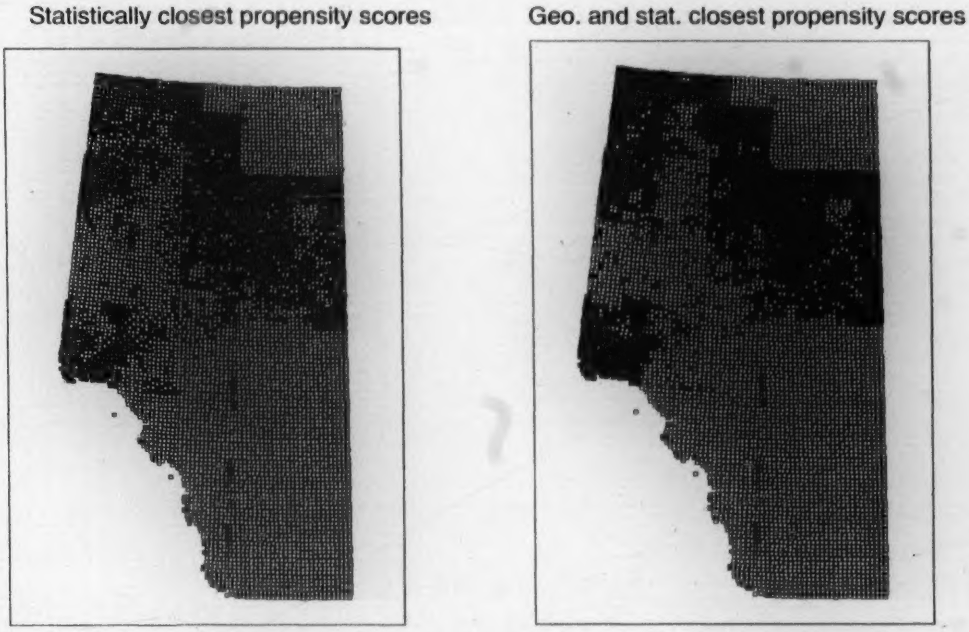
Predicted probabilities are obtained based on the output of the logit model given in Table 6.5. These predicted probabilities are the propensity scores. Out of the 3566 observations, 1167 are located in caribou habitat and 2399 observations are non-caribou habitat townships. Out of the 2399 observations located in non-caribou habitat, 1167 townships were selected as a control group using nearest neighbor matching or matching based on closest propensity score. For example if the propensity score of a given township is 0.88 then the algorithm will choose a propensity score equal to 0.88 or very close score and match it with the given township[45].

Two procedures are used for the matching process. The first procedure is matching based on the statistically closest propensity score and the second procedure is based on geographical proximity. The first method chooses a township from the study area and matches it to the township in a caribou area based on the statistically closest score. For the second method, a geographical location was imposed and the algorithm was forced to choose the statistically closest propensity score within the given geographical location. The PSAC area maps are used to impose geographical proximity. For example, for the Cold Lake caribou herd, the matching procedure was made to choose townships from region 6 of the PSAC area. For the Caribou Mountain herd, the algorithm was made to choose control group townships from region 7 of the PSAC map and so on. Maps of caribou habitat and the corresponding matched control groups (non-caribou habitat) for the whole study area and for selected caribou herds are given in Figures 6.9 to 6.12. Green areas show caribou habitat and red areas are matched non-caribou habitat.

---

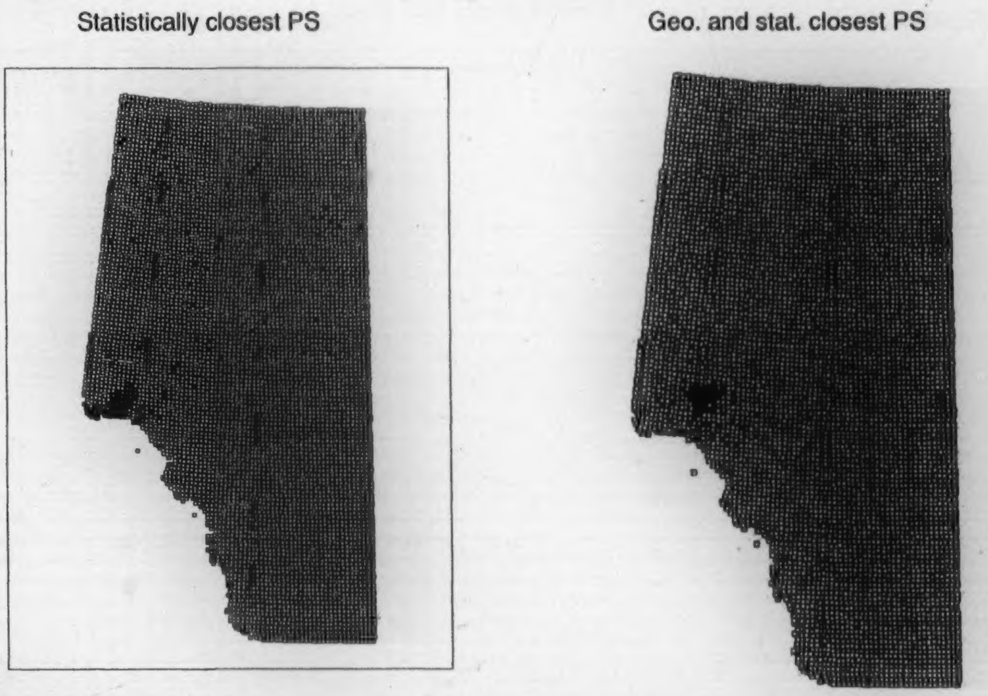
[45] A program on excel spreadsheet developed by Thomas Love (2004) at Case Western Reserve University is used for matching the estimated propensity scores using nearest neighbor matching.

Figure 6.9 Map of matched caribou and non-caribou habitat: all study areas


Statistically closest propensity scores

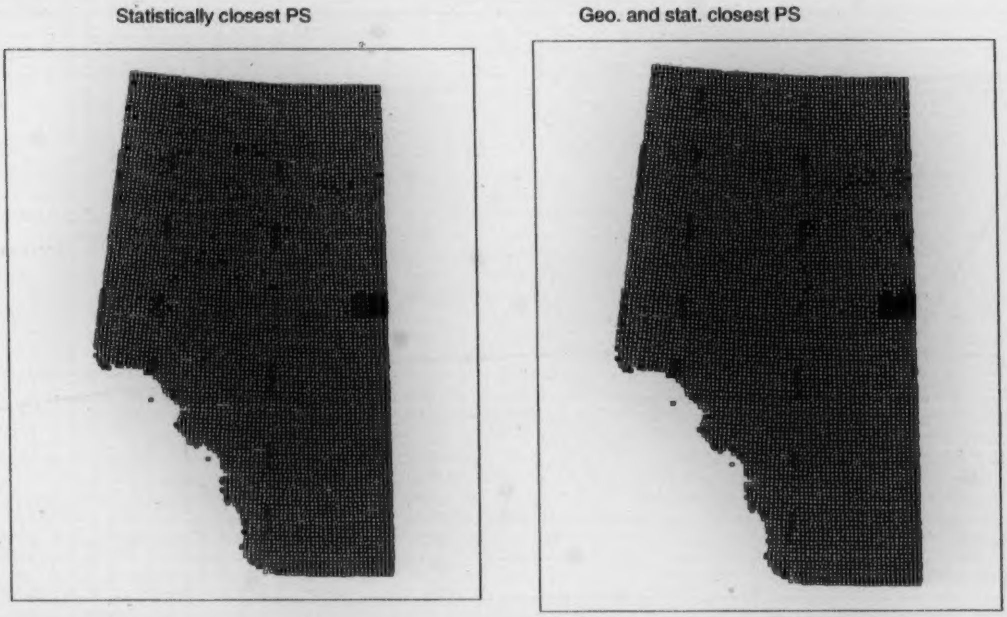
Geo. and stat. closest propensity scores

Figure 6.10 Map of matched caribou and non-caribou habitat: A La Peche Herd


Statistically closest PS

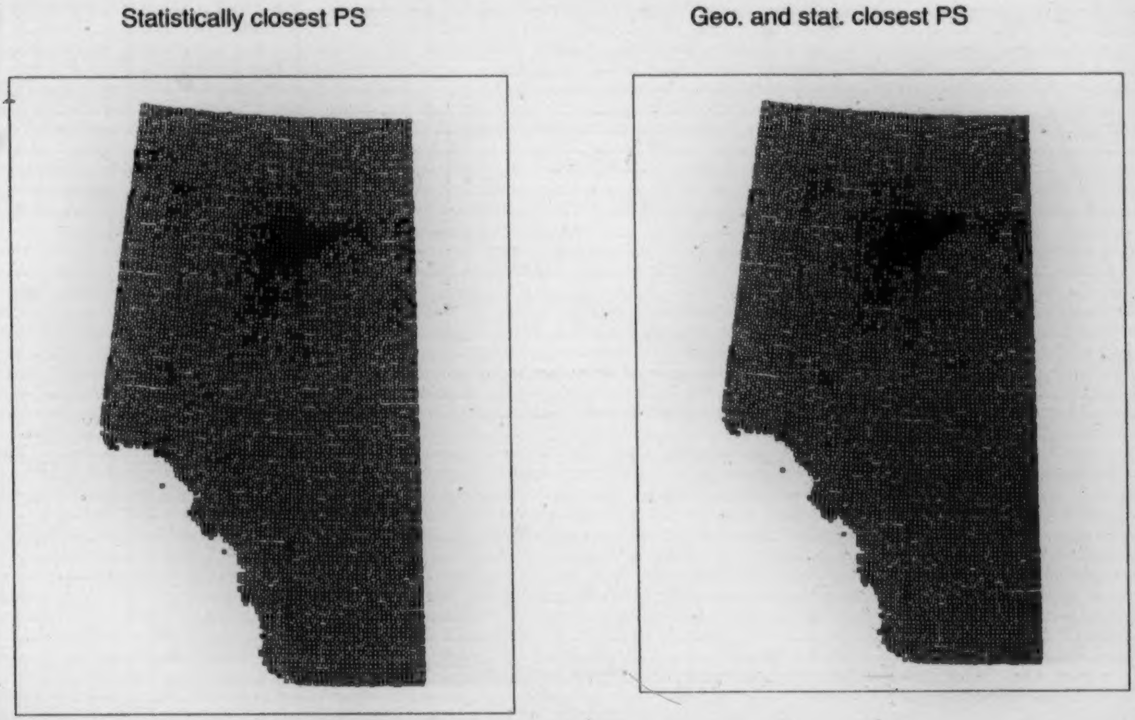
Geo. and stat. closest PS

Figure 6.11 Map of matched caribou and non-caribou habitat: Cold Lake Herd


Statistically closest PS

Geo. and stat. closest PS

Figure 6.12 Map of matched caribou and non-caribou habitat: Red Earth


Statistically closest PS

Geo. and stat. closest PS

Figures 6.9 to 6.12 show that when the statistically closest propensity score is used the matching algorithm tends to choose non-caribou habitat townships from all over the study area. The geographical variables included in the logit model do not seem to have a significant influence in the matching process. If matching based on closer geographical location is used, the non-caribou habitat clusters around the caribou habitat. In this case matching is performed at the expense of losing the statistical significance of the control groups that had closer propensity scores. The reason for introducing matching using geographical location is based on the assumption that geographical proximity can capture different unobservable or unavailable variables that determine drilling.

Counter checking the performance of the matching algorithm is an important step in the process of propensity score matching method. T-statistics for mean differences in the covariates are used to test if the matching algorithm has reasonably matched the two groups. Results for the statistical test of mean differences of the covariates are given in Table 6.6. For example, in column one of the results section the average difference for the cumulative wells between caribou habitat and matched non-caribou habitat is -0.022 and this parameter is not statistically significant (t = -1.55). Percentage of forest cover and its square, muskeg, and the longitude variables have, however, statistically significant means[46]. Results for the second procedure where townships are matched based on geographical location are similar to the first procedure. The only exception is that the mean difference for average *success rate* becomes significant and the significance level has increased for most of the variables.

---

[46] Matching results excluding forest cover and muskeg variables are discussed in the next section.

Table 6.6 Statistical tests of mean difference of the covariates for the matched groups

	Statistically Matched	Geo and Stat Matched
Variable	Mean diff.	Mean diff.
Cumulative Wells	-.022	-.025
	(-1.55)	(-1.77)
Cum. Wells Squared	-.036	-.030
	(-0.50)	(-0.91)
Avg. Success Rate	-.027	-.059
	(-1.73)	(-2.16)
Forest Cover (%)	-.200	-.150
	(-17.10)	(-12.91)
Forest Cover Squared	-.180	-.177
	(-13.36)	(-11.01)
Water Body Cover (%)	.000	.001
	(0.84)	(1.14)
Muskeg Cover (%)	.145	.1526
	(15.06)	(13.73)
Human Disturbance (%)	-.008	-.003
	(-.77)	(-1.34)
Roads	-.301	-.006
	(-.68)	(-1.54)
Roads Squared	2.02	.990
	(1.09)	(1.44)
Area of Township	-.032	-.388
	(-.09)	(-0.98)
Easting (Latitude)	-.006	-0.005
	(-.72)	(-1.13)
Northing (Longitude)	.004	0.011
	(3.55)	(4.09)

Note: t-statistics are in parenthesis.

Given the large number of covariates one can conclude that the covariates are reasonably matched. Within the different caribou herds there are significant ecological and geographical differences, and given that spatial and temporal exploration of oil and gas activities are heterogeneous, finding an exact match between townships in caribou habitat and non-caribou habitat is not an easy task. For example, mean differences of the covariates for the Caribou Mountain and its corresponding non-Caribou Mountain herd townships are in most cases statistically significant while mean difference of the covariates for the Cold Lake are not statistically significant. The results reported in Table 6.6 are the average for all herds.

Figures 6.13 to 6.18 show graphical representations of average drilling densities for caribou habitat and matched non-caribou habitat for the whole study area and for selected caribou ranges. The graphs show matching based on the two procedures discussed above. Figure 6.13 shows that average drilling is always higher in caribou habitat even after geographical restrictions were made on the matching procedure. Figure 6.14 shows that average drilling was clearly higher in caribou habitat after the Caribou Mountain and Bistcho herds were excluded from the analysis. Figure 6.15 shows average drilling after the Cold Lake herd is excluded in addition to the Caribou Mountains and Bistcho herds. Drilling in caribou habitat is still higher after Cold Lake is excluded. Analysis excluding these herds is done as these regions appear to be outliers in terms of drilling activity.

Figure 6.13 Average drilling densities for the matched caribou habitat and non-caribou habitat: All study area

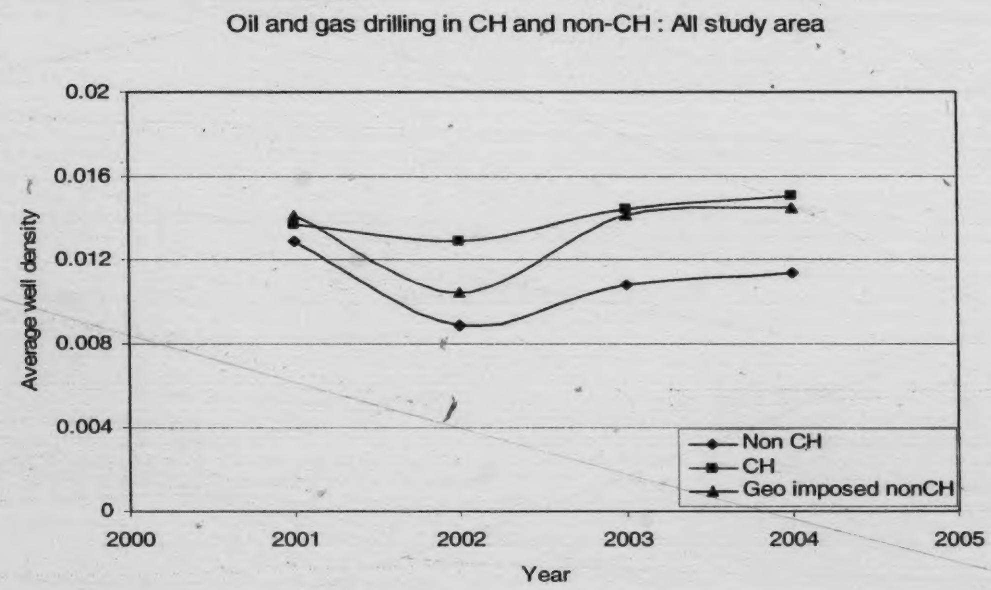


Figure 6.14 Average drilling densities excluding Caribou Mountain and Bistcho herds

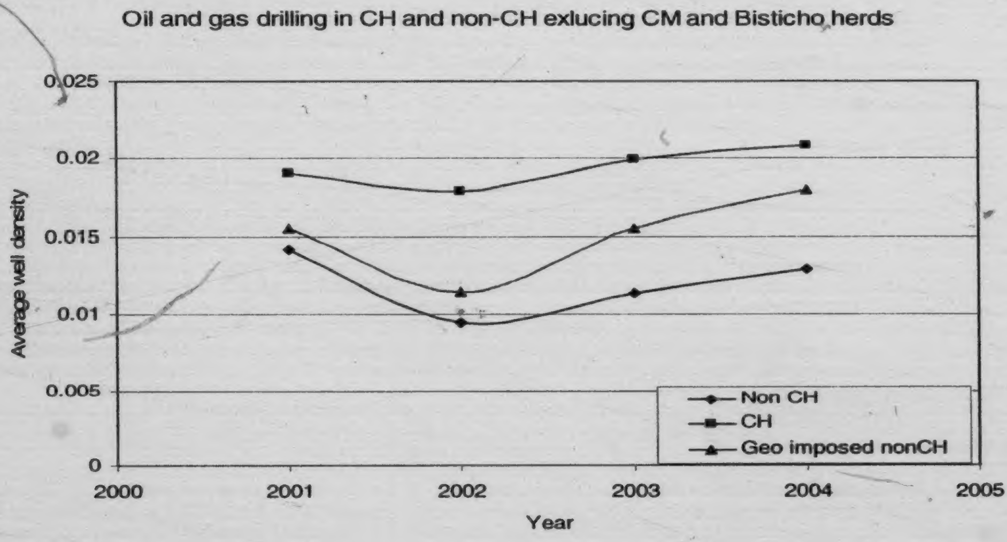


Figure 6.15 Average drilling densities excluding Caribou Mountain , Bistcho, and Cold Lake herds

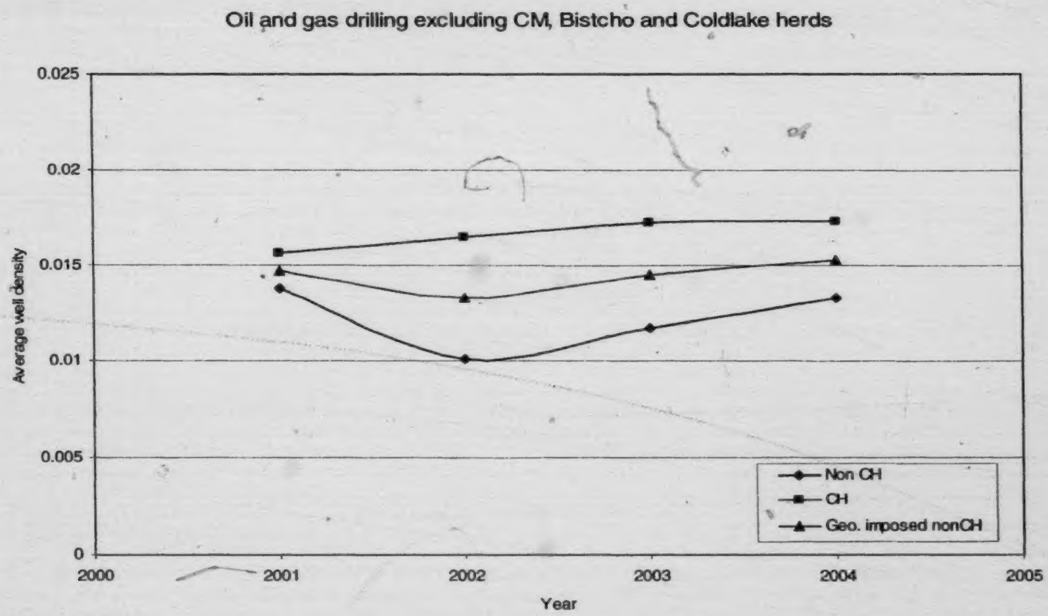


Figure 6.16 Average well densities for the matched caribou and non –caribou habitat: A La Peche

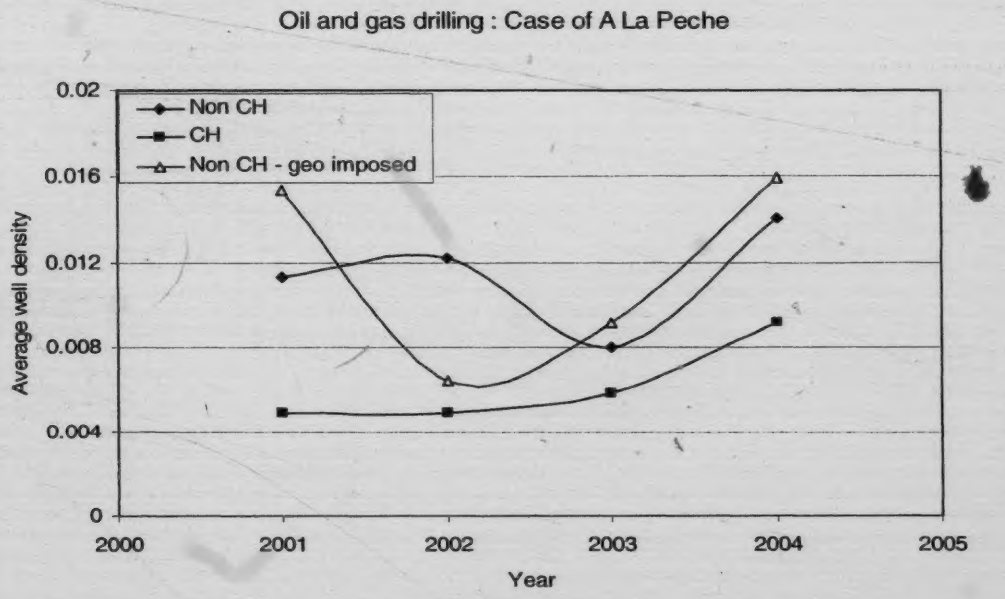


Figure 6.17 Average drilling densities for the matched caribou and non-caribou habitat: Red Earth

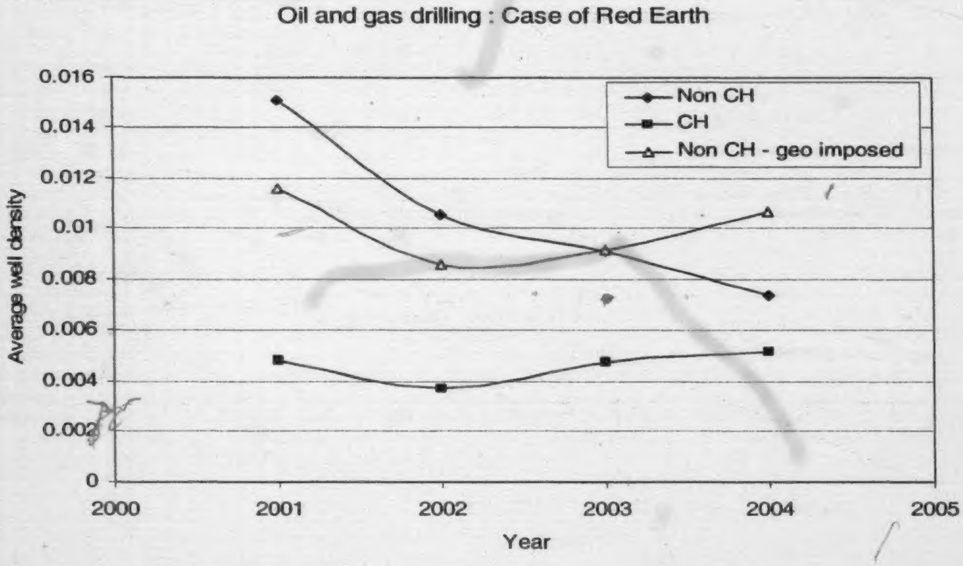
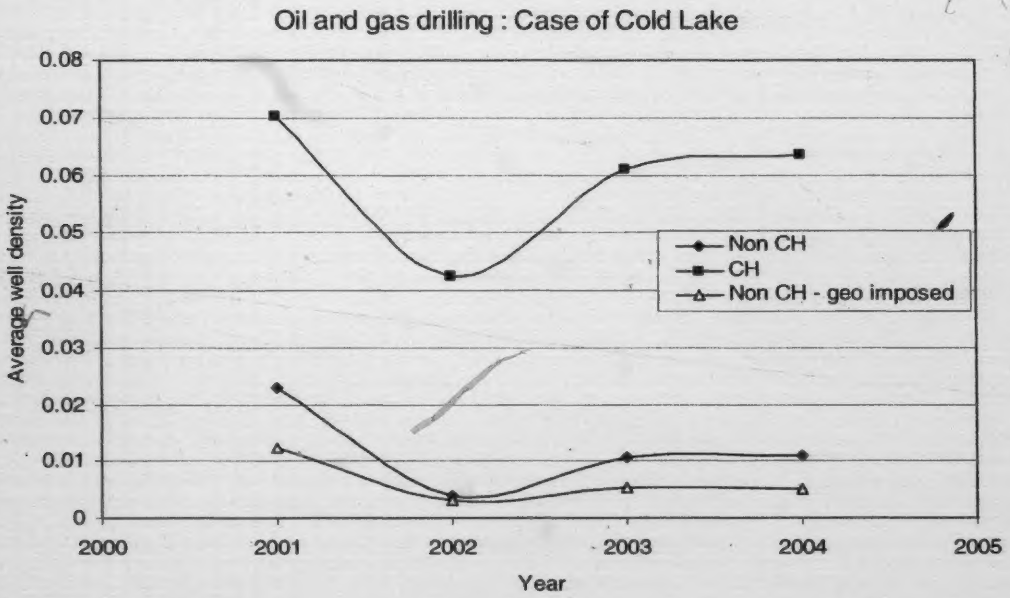


Figure 6.18 Average drilling densities for the matched caribou and non-caribou habitat: Cold Lake



Figures 6.16 to 6.18 show a comparison of average drilling for A Le Peche, Red Earth, and Cold Lake herds and their corresponding matched non-caribou habitat. Figures 6.16 and 6.18 show that average drilling in A Le Peche and Red Earth are lower than the matched non-caribou habitats. Similar results are obtained when matching was performed using the statistical and geographical procedures. However, for the case of Cold Lake average drilling was higher through out the study period. These results explain the spatial and temporal variation of well densities among the different herds through out the study period. Having analyzed the average difference of well densities between caribou and non-caribou habitat, the next logical step is to test the statistical significance of these differences. Results for the statistical tests of the mean difference for the whole study area and for selected caribou herds are given in Tables 6.6 and 6.7 below.

Table 6.7 Mean difference of well densities for the matched groups: all herds

	All study area		Study area excluding CM and Bistcho herds	
	Statistically Matched	Geo - Stat Matched	Statistically Matched	Geo - Stat Matched
Year	Mean diff.	Mean diff.	Mean diff.	Mean diff.
2001	.001 (0.40)	-.001 (-0.19)	.004 (1.32)	.003 (1.12)
2002	.004 (2.13)	.002 (1.107)	.007 (2.88)	.002 (2.08)
2003	.004 (1.64)	.000 (0.09)	.007 (2.47)	.001 (2.01)
2004	.004 (1.60)	.001 (0.21)	.007 (2.35)	.005 (1.98)

Note: t-statistics are in parenthesis

Table 6.7 presents the results for the statistical significance of the average difference of well densities between caribou habitat and non-caribou habitat. The results are repeated in the fourth and fifth columns after Caribou Mountain and Bistcho herds are excluded from the sample. A positive coefficient shows that average drilling is higher in caribou habitat for the given year. Except for the year 2001 (Geographically and statistically matched groups); the results show that average drilling in caribou habitat is higher than their corresponding non-caribou habitat in most of the cases. The results are not statistically significant in most cases when the whole study area is chosen. However, statistically significant differences are observed after Caribou Mountain and Bistcho herds are excluded from the analysis. For example, for the years between 2002 and 2004, the parameters are now statistically significant for both the statistically and geographically matched groups.

Analysis of individual caribou herds is more informative than the average results for all herds. Case studies of statistical significance of mean differences of drilling densities for the Cold Lake and Slave Lake herds are given in Table 6.8. The results show that in all cases the average drilling in the Cold Lake and Slave Lake herds is higher than their corresponding non-caribou habitat. For the Cold Lake herd the coefficients are statistically significant for the specified study period. For the case of Slave Lake the coefficients are all significant for all the years except for the year 2001. For the geographically and statistically matched groups the statistical significance decreases and in the years 2001 and 2002 it becomes insignificant.

Table 6.8 Mean difference of well densities for the matched groups: Cold Lake and Slave Lake herds

	Cold Lake		Slave Lake	
	Statistically Matched	Geo - Stat Matched	Statistically Matched	Geo - Stat Matched
Year	Mean diff.	Mean diff.	Mean diff.	Mean diff.
2001	.047 (1.60)	.058 (1.98)	.008 (1.69)	.005 (0.51)
2002	.039 (3.01)	.039 (3.08)	.007 (2.73)	.005 (1.21)
2003	.051 (2.40)	.056 (2.67)	.015 (3.92)	.012 (2.77)
2004	.053 (2.72)	.059 (3.04)	.033 (2.60)	.033 (2.55)

Note: t-statistics are in parenthesis

One difficulty with the propensity score matching is that the forest cover and muskeg variables are difficult to match between caribou and non-caribou habitat. This could be attributed to the unique nature of caribou habitat. Since the primary interest of this analysis is matching based on energy sector factors, the next section presents the results of an analysis excluding forest cover and muskeg variables from the logit model.

6.7.4 Propensity Score Matching Results – Part II

The results in Table 6.6 show that statistically significant differences are observed between caribou and matched non-caribou habitat in percentage of forest cover and muskeg. Another model was estimated excluding these variables. Theoretically, excluding these variables could make sense if we assume that differences in forest cover and/or muskeg between the two habitats does not affect the decision to explore in either habitat. A summary of the results excluding these variables are given in Table 6.9.

Table 6.9 Results for mean difference of covariates and well densities: excluding forest cover and muskeg.

	Average of all herds	All herds excluding CM and Bistcho
Variable	Mean diff.	Mean diff.
Cumulative Wells	-.021	-.025
	(-1.57)	(-1.97)
Cum. Wells Squared	-.036	-.030
	(-0.50)	(-0.41)
Avg. Success Rate	-.027	-.059
	(-1.43)	(-3.16)
Water Body Cover (%)	.001	.002
	(0.34)	(1.14)
Human Disturbance (%)	-.002	-.004
	(-.70)	(-1.34)
Roads	-.341	-.004
	(-.88)	(-1.34)
Roads Squared	1.02	.980
	(.99)	(1.44)
Area of Township	-.032	-.355
	(-.08)	(-0.96)
Easting (Latitude)	-.005	-0.004
	(-.82)	(-1.03)
Northing (Longitude)	.003	0.007
	(4.25)	(4.39)
Drilling density 2001	-.004	.001
	(-1.54)	(.52)
Drilling Density 2002	.001	.006
	(.27)	(2.15)
Drilling Density 2003	.001	0.004
	(.46)	(2.03)
Drilling Density 2004	.002	0.004
	(.68)	(1.12)

Note: t-statistics in parenthesis

The results in Table 6.9 show that most of the covariates are now statistically matched. However, mean differences in drilling densities are not statistically significant for all the study period. It is even negative for the year 2001. Once Caribou Mountain and Bistcho are excluded from the match, average well densities become significant for the years 2002 and 2003. Even though the matching algorithm has performed better once the forest cover and muskeg were excluded, the mean difference between well densities did not significantly change.

6.8 Summary of results and hypothesis tests

Based on the results obtained using the multivariate regression model, the difference in difference method, and the propensity score matching method, we can conclude that the results are generally consistent regardless of the method used. Average drilling in caribou habitat appears to be higher than non-caribou habitat in most of the cases. There are some exceptions. For example, in some cases negative coefficients are observed for the year 2001 when a comparison is made using the propensity score method. However, this difference was not statistically significant. Regarding the statistical significance of the coefficients, most of the coefficients obtained using the multivariate regression and the difference in difference methods are significant at a 5% level of significance. However, the coefficients obtained using the propensity score method are not statistically significant when the entire study area is considered. Most of the coefficients become significant after Caribou Mountain and Bistcho herds are excluded from the sample. It is also observed that mixed results are obtained for the analysis based on individual herds.

The aim of this analysis was to test the hypothesis that the energy sector anticipates new regulations from the government and may increase its exploration activities before the regulations are implemented. Two historical events were considered to test this hypothesis. The first event was the year 1991 when the provincial government placed woodland caribou on the 'Red List' as a species at risk of extinction. It was hypothesized that this precautionary information could make the energy sector engage in more exploration activities in a caribou area before a restriction was placed that prohibited drilling in these areas. Using the difference in difference method drilling activities in caribou and non-caribou habitat was observed before and after 1991. The coefficient for caribou habitat after 1991 was positive and significant (0.0017). Given a positive result we can conclude that the energy sector was not slowing down its exploration activities after this precautionary information was released. Hence the hypothesis could not be rejected.

The second historical event was the year 2001 when the provincial government re-designated woodland caribou as a threatened species. This precautionary information was considered to test the hypothesis similar to the first case discussed above. Results from the difference in difference method and from the propensity score matching confirm that drilling activities were on average higher in caribou habitat immediately after the year 2001. These results are statistically significant for the difference in difference method and for the propensity score method after Caribou Mountain and Bistcho herds are excluded. Results from the individual herds show that after the year 2001 higher drilling activities were observed in the Cold Lake, Richardson, and WSAR herds. Based on these findings we cannot reject the hypothesis that this information would give a signal to the energy sector to engage in more exploration activities in caribou area before a restriction that prohibits drilling is implemented.

Further study is however needed to investigate if the energy sector was maintaining the integrity and supply of habitat using drilling techniques that permit its use by caribou. Moreover, further study is required to analyze if the statistically significant differences are also ecologically significant to affect caribou habitat. For example, the results from the multivariate regression model show that on average about 118 more wells per year are drilled in caribou habitat compared with non-caribou habitat[47] after the other explanatory variables are controlled. Throughout the study period the number of wells drilled in non-caribou habitat (on the specified study area) range between 700 and 5500 wells and number of wells in caribou area vary between 140 and 1600 wells. In sum, the positive and statistically significant coefficients of drilling activities in caribou habitat show that there appear to be higher rates of exploration of oil and gas activities in caribou habitat. These findings show that management agencies should consider the anticipatory effect of their regulations when designing strategies for endangered species conservation and recovery. Specific management or policy implications of the study and recommended strategies are discussed in the next section.

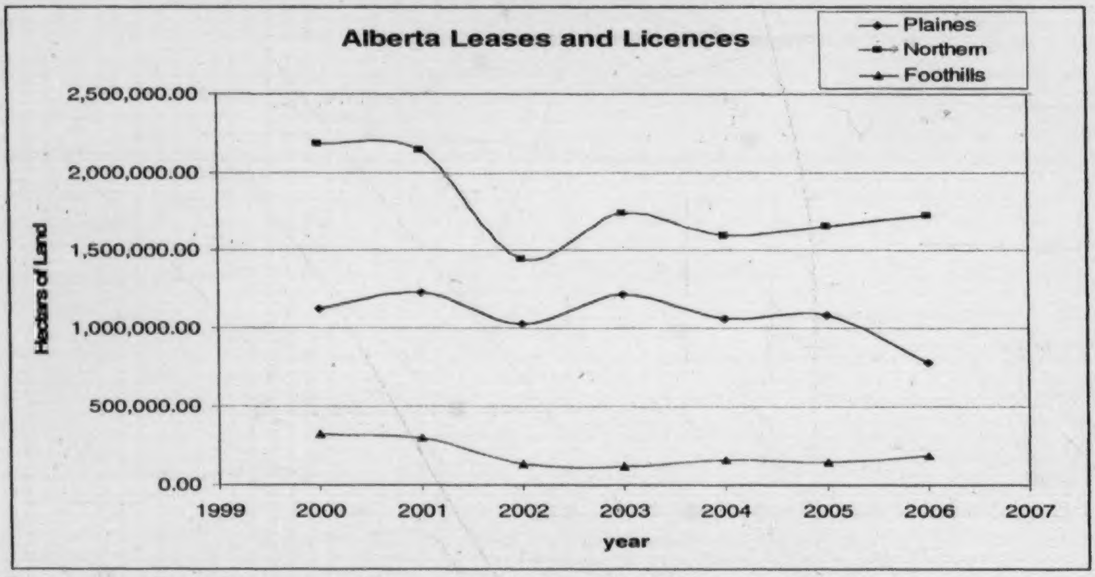
---

[47] The average area of a township is 92 sq.km and there are 1167 CH townships. Multiplying these figures by the coefficient gives 118 wells. i.e. $92km^2 \times 0.0011 wells/km^2 \times 1167 townships = 118 wells$

6.9 Management implications

It is now widely cited in the literature that caribou populations are declining and the most detrimental factor for caribou population dynamics is the functional loss of habitat due to reduced use of quality habitat in proximity of industrial activities (Alberta Woodland Caribou Recovery Team 2005, Dyer et al 2001, Dzus 2001). The findings from the present study show that considerable oil and gas activities are observed in caribou habitat especially in the last few years of the study period (2001 to 2004). Moreover, Figure 6.19 below shows that trends in hectares of land leased and licensed to the energy sector are increasing in the Northern and Foothills regions. Specifically, the figure shows an increase in the trend of land leased and licensed in the Northern region in the last few years (2004 to 2006). The same is true for the Foothills region. Given that petroleum and natural gas leases last for five years and the fact that petroleum and natural gas licenses are issued for an initial term of four years in the Northern region and five years in the Foothills region, significant drilling activities are expected in these regions for the next several years. All the caribou habitat in Alberta is located either in the northern or foothills regions. These facts show the need for long-term management of caribou habitat to maintain a viable population.

Figure 6.19 Hectares of land leased and licensed in Alberta



Data source: Alberta Department of Energy: Petroleum and Natural Gas Sales Statistics

The results from the study show that once caribou is listed as endangered or threatened there appears to be no trend of decreasing oil and gas exploration activities in caribou habitat. In some cases like the Cold Lake and WSAR, and recently for herds such as Richardson, exploration activities appear to be significantly higher than the corresponding non-caribou habitat.

The results from this study show that heterogeneous exploration activities are observed over time and among the different herd regions. In some ranges no major exploration activities were observed in the past few years but recently the trend has changed and significant activities are observed. In other cases there are no major exploration activities on caribou habitat. Based on these facts, minimizing the impact of future development on caribou ranges and recovery of the existing industrial footprints should be implemented based on range specific management plans for all the herds in Alberta. Even though this study was focused on oil and gas exploration activities, caribou habitat is not only disturbed by the energy sector but also by forestry and other sectors. Hence, coordination of activities of the energy sector and the timber industry are very important in minimizing industrial footprints in caribou range habitats. The government's commitment to sustainable resource and environmental management plans which was formed in 1999 by the Department of Energy, Sustainable Resource Development and Alberta Environment is an ideal institution to take the initiative and implement the coordination activities.

Chapter 7 Conclusions

The thesis was organized into three parts. The first part has provided a modeling approach and empirical results regarding the energy sector spatial and temporal exploration activities in Alberta. The specification and estimation of the oil and gas exploration model was built by incorporating economic and geologic variables that determine exploration activities. Economic theory issues related to uncertainties, exploration through learning, and clustering and depletion effects of reserves were also incorporated in building the model. Estimating a spatial panel data models is a complex task that requires combining econometrics and geographic information systems (GIS). Due to the spatial nature of the data, different spatial econometric specifications were used to estimate the model. These include the spatial fixed effects, time-period fixed effects, and the spatial and time-period fixed effects of the spatial lag and spatial error models. Moreover, regional specifications of the model were taken into account during model development and estimation processes. The major findings of the first part include evidences of clustering of deposits, depletion effect of resources, and learning from exploration.

The main limitations of the exploration model include the following: First, data on oil and gas reserves are not available on a yearly basis. Hence, cumulative number of wells was used as a proxy variable to capture reserves. Second, theoretically the spatial econometric model can contain both a spatially lagged dependent variable and spatially auto-correlated errors simultaneously. However, the models estimated in this study were confined to either spatial lag or spatial error models. A model that combines both terms would give more robust results. In practice, such models are rarely used. However, efforts are underway to incorporate spatial lag and spatial error models simultaneously (Aneslin 2004).

The first part of the study can be extended in at least two ways. First, including the spatial lags of some of the explanatory variables would give more insights. For example, the present study included temporal lag of *success rate* as one of the explanatory variables. Adding a spatial lag of *success rate* as an additional explanatory variable could be an alternative modeling approach. In this case, the model would be built based on the assumption that firms would collect information on the success rate of wells not only from a given township but also from neighborhood townships. The same is true for the cumulative number of wells. Second, an alternative modeling approach is to model oil and gas exploration based on major stratigraphic zones instead of regions. In Alberta oil and gas leases and license are allocated on stratigraphic interval bases, which indicate that drilling effort could differ among different horizons.

The second part of the thesis was on forecasting oil and gas drilling efforts. The forecasting model was based on the spatial lag model. Understanding future energy sector spatial exploration activities is an important input to land use planning and the cumulative effects management initiatives taken by the government of Alberta[48]. Forecasting exercises are mostly practiced in the time-series literature and are implemented on aggregated national basis. In the present study, forecasts are made at a township level. The advantage of forecasting at a smaller scale level is that policy decisions concerning the impacts of industrial activities and land-use planning are mostly implemented on a specific region.

---

[48] The Government of Alberta is taking an initiative to develop and implement land use planning in order to address a wide range of land management issues. More information is found at www.landuse.gov.ab.ca.

The main limitation of the forecasting model is that the model does not have an upper limit on the maximum number of wells that can be drilled in a given township. The negative coefficient for the cumulative number of wells ensures that at some point in time the curve will be downward slopping. However, the upper limit of the curve and the maximum number of wells allowed in a township could be different. According to the EUB regulations, the spacing unit for oil wells is normally one well per quarter section of land and for gas wells it is one well per section of land[49]. Even though these targets are not reached for the specified forecasting period (2005 to 2020), forecasts beyond this period could exceed the maximum limit. The other limitation is that the model would tend to forecast wells in all townships regardless of whether there is evidence of reserves or not. Even though some clustering tendency is observed in the forecasted wells, further study is required to develop a model that takes into account clustering in a rigorous way. One suggested approached is the inclusion of spatial lag of cumulative wells as one of the explanatory variables. This is, however, an empirical question that should be tested using the data.

---

[49] There are some exceptions to these rules especially in areas where there are high reserves of oil and gas (needs approval from EUB). Definitions for sections and quarter of a section in a township are explained in Appendix 1.1.

The main focus of the third part of the thesis was to test the hypothesis that the energy sector would anticipate new regulations that protect caribou habitat and increase their exploration activities before the new regulation is implemented. On overage, higher drilling are observed in caribou habitat than in non-caribou habitat. This suggests that the energy sector tends to increase exploration activities after the information regarding the endangerment of caribou species was released. However, the results are mixed for herd-specific models. The fact that oil and gas resources and wildlife habitat in Alberta are publicly owned makes this study different from previous studies in the US. Unlike previous studies, testing the hypothesis using different statistical tools has helped the study to obtain robust results. Applying these methods to the present study is, however, not with out limitations. For example, the main limitation of the multivariate regression and the difference in difference method is that the models could suffer from functional form misspecifications. Moreover, matching based on geological and ecological covariates is not an easy task. The matching results show significant differences among the variables such as forest cover and muskeg. The other limitation of the matching procedure is luck of data on reserves of oil and gas at a township level in the specified time period. Instead a proxy variable represented by cumulative number of wells is used. Matching based on reserves data would give more robust results because reserves of oil and gas are the main factors that determine exploration activities. Regarding the matching algorithm, only the nearest neighbor matching algorithm is used. Several other options such as kernel or caliper matching could also be used. The nearest neighbor algorithm is however recommended if there are enough control groups to choose from the sample. Further research is required to address the above mentioned limitations.

The present study has made substantial contribution to our understanding of the oil and gas exploration activities in terms of the factors that affect exploration, future trend of exploration on the landscape, and the energy sector anticipation of new environmental regulations in Alberta. Nevertheless, an interesting future research in this area could be to specifically select a smaller grid of land and test the hypothesis of the energy sector's anticipation of new environmental regulation in wildlife habitat in terms of the statistical and ecological significance of exploration activities in the specified habitat. Selecting a smaller area of land has the advantage of capturing all the relevant variables and avoiding biases introduced due to averaging out of exploration activities. Moreover, selecting a smaller area has the advantage of organizing the data in to smaller resolution such as sections or sub-sections of the land instead of bigger townships.

References

Abadie, A. 2005 "Semi-parametric Difference-in-Differences Estimators." *Review of Economic Studies,* Vol. 72(1): 1-19.

Alberta Woodland Caribou Recovery Team. 2005 "Alberta Woodland Caribou Recovery Plan 2004/05-2013/14.", Alberta Sustainable Resource Development, Fish and Wildlife Division, Alberta Species at Risk Recovery Plan No. 4. Edmonton, AB.

Anselin, L. 2002, "Under the hood: Issues in the specification and interpretation of spatial regression models." *Agricultural Economics*, 27: 247-267.

Anselin, L., 1988, *Spatial Econometrics: Methods and Models.* Dordrecht, Netherlands: Kluwer.

Anselin, L., 1995, "Local indicators of spatial association – LISA", *Geographical Analysis*, Vol. 27 (2).

Arrow, K. J. and Chang, S. 1978, "Optimal Pricing, Use and Exploration of Uncertain Natural Resource Stocks." *Technical Report No. 31*, Harvard University.

Attanasi, Emil D. 1979, "The Nature of Firm Expectations in Petroleum Exploration." *Land Economics* (55):299-312.

Baltagi, B.H. and Li, D. 2004 "Prediction in Panel Data Model with Spatial Correlation.", in '*Advances in Spatial Econometrics: Methodology, Tools and Applications*'. Anselin, L. , Florax, R.J.G.M. and Rey, S. H. (Germany), Springer-Verlag:283 – 295.

Baltagi, B.H. and Li, D. 2006 "Prediction in the Panel Data Model with Spatial Correlation: the Case of Liquor." *Spatial Economics Analysis*, Vol. 1(2): 175 – 185.

Cairns R. D and Quyen, N. 1998 "Optimal Exploration for and Exploitation of Heterogeneous Mineral Deposits" *Journal of Environmental Economics and Management*, 35: 164 – 189.

Cairns R. D (1990) "The economics of exploration for non-renewable resources" *Journal of Economic Surveys* Vol 4 (4): 361 – 395.

Cairns R. D (1994) " On Gray's rule and the stylized facts of non-renewable resources." *Journal of Economics Issues*. Vol 28, pp. 777 – 798.

Caliendo, M. and Kopeinig, S. 2005 "Some Practical Guidance for the Implementation of Propensity Score Matching", Institute for the Study of Labor (IZA*), IZA Discussion Papers* #1588.

CAPP (Canadian Association of Petroleum Producers), 2005 "The official web site of Canadian Association of Petroleum Producers" www.capp.ca.

Card, D. and Krueger, A. 1994 "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *American Economic Review*, 84(4): 772 – 93.

Chomitz Kenneth, M. and Gray David A. 1996 "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize" *The World Bank Economic Review* Vol.10 (3): 487-512.

Dahl, C. A., and Duggan, T. 1998 "Survey of Price Elasticities from Economic Exploration Models of US Oil and Gas Supply," *Journal of Energy Finance and Development,* 3(2): 129-169.

Dehejia, R. and Wahba, S. 1999. "Causal Effects in Non Experimental Studies: reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association,* Vol.94.

Dehejia, R., and Wahba, S. 2002 "Propensity Score-matching Methods for Non Experimental Causal Studies" *The Review of Economics and Statistics*, 84 (1), 151-161.

Deshmukh, S.D. and Pliska, S.R. 1980, "Optimal Consumption and Exploration of Non-renewable Resources under Uncertainty." *Econometrica*, (48): 177 – 200.

Devarajan, S. and Fisher, A. 1981, "Hotelling's Economics of Exhaustible Resources: Fifty Years Later" *Journal of Economic Literature*, Vol. 19 (1) : 65-73.

DOE (Alberta Department of Energy) 2005 "The official website of Alberta Department of Energy" http://www.energy.gov.ab.ca

Dyer, S.J., J.P. O'Neill, S.M. Wasel, and S. Boutin. 2001 "Avoidance of Industrial Development by Woodland Caribou." *Journal of Wildlife Management* 65(3): 531-542.

Dzus, E. 2001. "Status of the Woodland Caribou in Alberta." Alberta Environment, Fisheries and Wildlife Management Division, and Alberta Conservation Association, *Wildlife Status Report No. 30*, Edmonton, AB.

East Side Range Planning Team, 2005, "*East Side of the Athabasca River Range Plan*" Boreal Caribou Committee.

Elhorst, J. Paul, 2003. "Unconditional maximum likelihood estimation of dynamic models for spatial panels," *Research Report 03C27*, University of Groningen, Research Institute SOM (Systems, Organizations and Management).

Elhorst, J.Paul, 2005. "Models for Dynamic Panels in Space and Time - an Application to Regional Unemployment in the EU," *ERSA conference papers, ersa05p81*, European Regional Science Association.

EUB (Alberta Energy and Utilities Board) 2004 "Alberta's Reserves and Supply/Demand Outlook 2004-2013" *Statistical Series (ST): 2004 -98*, Calgary.

EUB (Alberta Energy and Utilities Board) 2005 "Alberta's Reserves and Supply/Demand Outlook 2005-2013" *Statistical Series (ST): 2005 -98*, Calgary.

Farzin, Y. Hossein (2001) "The Impact of Oil Price on Additions to U.S. Proven Reserves" *Resource and Energy Economics*, Vol. 23 (3): 271 – 292.

Ferraro, P.J., McIntosh, C. and M. Ospina, M. 2007 "The Effectiveness of Listing under the U.S. Endangered Species Act: An econometric analysis using matching methods" *Journal of Environmental Economics and Management,* 54(3): 245 – 261.

Fisher, F.M., 1964. "*Supply and Costs in the US Petroleum Industry.*" John Hopkins University Press, Baltimore, MD.

Forbes, K.E. and E.M. Zampelli, 2000 "Technology and the exploratory success rate in the US offshore" *The Energy Journal* 21:109-120.

Gaudet, G., Moreaux, M., Salant , S.W., 2001 "Inter temporal Depletion of Resources Sites by Spatially Distributed Users", *American Economic Review,* 91:1149 – 59.

Green, W. 2000. "*Econometric Analysis,*" 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

Greenstone, Michael and Gayer, Ted, 2007 "Quasi-Experimental and Experimental Approaches to Environmental Economics." *RFF Discussion Paper*, 07 – 22.

Greenstone, Michael. 2002. "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures." *Journal of Political Economy* 110: 1175-1219.

Greenstone, Michael. 2004. "Did the Clean Air Act Amendments Cause the Remarkable Decline in Sulfur Dioxide Concentrations?" *Journal of Environmental Economics and Management* 47(3):585 – 611.

Hallstrom, D. G., & Smith, V. K. 2005 "Market Responses to Hurricanes." *Journal of Environmental Economics & Management,* 50(3), 541-550.

Hauer, G., Steve Cumming, Fiona Schmiegelow, Wiktor Adamowicz, Marian Weber, Robert Jagodzinski. 2007 "Tradeoffs between forestry resource and conservation values under alternate forest policy regimes: A spatial analysis of the western Canadian boreal plains". *Rural Economy Staff Papers, 07-03*, 52p.

Heal, G. 1979. "Uncertainty and the Optimal Supply Policy for an Exhaustible Resource." In *Advances in the Economics of Energy and Resources, Vol. 2.* Jai Press, Greenwich.

Heckman, J. , Hidehiko , H. and Todd, P. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies* 64(4):605–54.

Hendricks, P. and Porter, R.H. 1996 "The Timing and Incidence of Exploratory Drilling on Off Shore Wildcat Tracts" *The American Economic Review*, Vol. 86 (3): 388 – 407.

Hotelling, H. 1931 "The Economics of Exhaustible Resources" *Journal of Political Economy*, 39(2): 137 – 175.

Hubbert M. King, 1956 "Nuclear Energy and Fossil Fuels" in *Drilling and Production Practice*, pp. 7 – 25, American Petroleum Institute, Washington D.C.

Iledare ,O. O. and Pulsipher, A.G. 2001 "The Determinants of Oil and Gas Reserves Additions: Empirical Evidence from the U.S. Gulf of Mexico Outer Continental Shelf (OCS) Region, 1977-1998" *SEP 68604*, Society for Petroleum Engineers (SPE) and Center For Energy Studies, Louisiana State University.

Iledare, O.O., 1995. "Simulating the effect of economic and policy incentives on natural gas drilling and gross reserve additions." *Resource and Energy Economics*, 17 (3):262-279.

Iledare, O.O. and Pulsipher, A.G., 1997 "The modeling of the petroleum exploration and extraction process for policy analysis: a case study of the Louisiana onshore region." , *Pac. Asian J. Energy* 7 (1), pp. 21–28.

Iledare, O.O., and. Pulsipher, A.G, 1999. "Sources of Change in Productivity of Petroleum Exploration and Development in Onshore Louisiana the US, 1977-1994." *Energy Economics*, Vol.21(3): 261 – 271.

Imbens, G.W. and Angrist, J.D, 1994. "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, Vol. 62 (2): 467-75.

Kaufmann, R. K. and Cleveland, C. J. 2001 "Oil Production in the lower 48 states: Economic, Geological and Institutional determinants. *The Energy Journal*, 22 (1): 27-49.

Kemp Alexander and Kasim Sola, 2003 "An Econometric Model of Oil and Gas Exploration, Development, and Production in the UK Continental Shelf: A Systems Approach" *The Energy Journal* 24 (2): 113 – 141.

Kemp Alexander and Kasim Sola, 2006 "A Regional Model of Oil and Gas Exploration in UK Continental Shelf.", Scottish *Journal of Political Economy,* 53 (2): 198 – 221.

Kholodilin, K.A, Siliverstovs, B., *Stefan* Kooths, S. 2007 "Dynamic panel data approach to the forecasting of the GDP of German Lander." German Institute for Economic Research, *Discussion Papers of DIW Berlin* # 664.

Kim, Chong Won, Phipps, Tim T. and Anselin Luc (2003) "Measuring the benefits of air quality improvement: a spatial hedonic approach" *Journal of Environmental Economics and Management* 45: 24 – 39.

Kolsrud Dag, 2004 "On Stochastic Simulation of Forward Looking Models" *Computational Economics*, 24: 159 – 183.

Krautkraemer, Jeffery A. 1998 "Non-renewable Resource Scarcity" *Journal of Economic Literature,* Vol. 36: 2065-2107.

Kunce Mitch, Gerking Shelby, and Morgan William 2004, "Environmental and Land Use Regulations in Non-renewable Resource Industries: Implications from the Wyoming Checkerboard" *Land Economics*, 80 (1): 76 – 94.

Lerch, I. and Mackay, J.A., 1999 "*Economic Risk in Hydrocarbon Exploration.*" Academic Press.

List, J. A., Millimet, D. L. and McHone, W. 2003 "The Unintended Disincentive in the Clean Air Act." *Advances in Economic Analysis and Policy*: Vol. 4 : Issue. 2, Article 2.

Livernois, John R 1988 " Estimates of Marginal Discovery Costs for Oil and Gas" *The Canadian Journal of Economics* Vol. 21 (2): 379 – 393.

Livernois, John R., and Uhler, Russell S. 1987 "Extraction Costs and the Economics of Non-renewable Resources." *Journal of Political Economy* 95: 195-202.

Longghi, S. and Nijkamp, P. 2007 "Forecasting Regional Labour Market Developments Under Spatial Autocorrelation", *International Regional Science Review*, Vol. 30 (2): 100 – 119.

Love Thomas, E. 2004 "Using Propensity Scores Effectively", Cleveland Chapter of the American Statistical Association Fall Workshop, Cleveland.

Lueck D. and Michael J. A. 2003 "Preemptive Habitat Destruction under the Endangered Species Act" *Journal of Law and Economics*, 46 (1): 27 – 60.

Lynch, Michael C. 2002 "Forecasting Oil Supply: Theory and Practice" *The Quarterly Review of Economics and Finance*" 42: 373 – 389.

MacAvoy, Paul W. and Pindyck, Robert S. 1973 "Alternative Regulatory Policies for Dealing with the Natural Gas Shortage" *The Bell Journal of Economics and Management Science*, Vol.4 No.2 : 454 – 498.

Managi, S., Opaluch, J. J., Jin, D. and Grigalunas, T. A. (2005). "Technological change and petroleum exploration in the Gulf of Mexico." *Energy Policy*, 33, pp. 619–32.

Margolis, M., Osgood, D.E., List, J.A 2005 "Does the Endangered Species Act Endanger Species? Evidence from a Natural Experiment", *NBER Working Papers 12777.*

Marr-Laing, T and Severson-Baker, C 1999 "Beyond *Eco-terrorism: the deeper issues affecting Alberta's oil patch*", Pembina Institute, Drayton Valleny, AB. (Available at www.pembina.org).

Mason , Charles F. 1985 "Learning from Exploration Information: The case of uranium" *Resources and Energy*, 7: 243 – 257.

Meyer, B. 1995 "Natural and Quasi-Natural Experiments in Economics," *Journal of Business and Economic Statistics,* (12): 151–162.

Meyer, B., Viscusi, W. and Durbin, D. 1995 "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment." *American Economic Review,* 85(3): 322-40.

Moroney John R. and Berg Douglas M. 1999 "An Integrated Model of Oil Production" *The Energy Journal* 20 (1): 105 – 124.

Mueller, M..J 1994 "Behavior of Non-renewable Natural Resource Firms Under Uncertainty." *Energy Economics*, 16 (1): 9 – 21.

Patton, M. and McErlean. S.A. 2003 "Spatial Effects within the Agricultural Land Market in Northern Ireland.", *Journal of Agricultural Economics* 55 (1): 127–133.

Patuelli, R.D., Griffith, M. T. and Nijkamp, P. (2006). "The Use of Spatial Filtering Techniques: The spatial and space-time structure of German unemployment data. *Tinbergen Institute Discussion Papers 06-049/3*, Tinbergen Institute.

Pesaran , Hashem M. 1990 "An Econometric Analysis of Exploration and Extraction of Oil in the U.K. Continental Shelf" *The Economic Journal*, Vol. 100 (401):367 – 390.

Pindyck, Robert S. and Rubinfeld, Daniel L. 1981 "*Econometric Models and Economic Forecasts*" second edition, Irwin McGraw-Hill, USA.

Pindyck, Robert S. 1978 "The Optimal Exploration and Production of Non-renewable Resources" *The Journal of Political Economy*, 86 (5): 841-861.

Pindyck, Robert S. 1980, "Uncertainty and Exhaustible Resource Markets", *Journal of Political Economy*, Vol. 88 pp. 1203 –25.

Quyen, Nguyen V. (1991) "Exhaustible Resources: A Theory of Exploration" *The Review of Economic Studies*, Vol 58 (4): 777 – 789.

Quyen, Nguyen V. 1988 "The Optimal Depletion and Exploration of Non-renewable Resource" *Econometrica*, 56: 1467 – 71.

Rosenbaum, P. and Rubin, D. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrica* 70(1):41–55.

Rosenbaum, P. and Rubin, D. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician* 39: 33-38.

Schneider, Richard 2002 "*Alternative Futures: Alberta's boreal forest at the crossroads*" Edmonton, Federation of Alberta Naturalists and Alberta Center for Boreal Research.

Schneider, Richard J. Stelfox, Brad J. Boutin, Stan. and Wasel, Shawn. 2003 "Managing the Cumulative Impacts of Land-uses in the Western Canadian Sedimentary Basin: A Modeling Approach." *Conservation Ecology*, 7 (1): Article 8.

Schneider, Richard and Dyer, Simon. 2006 "*Death by a Thousand Cuts: Impacts of in situ oil sands development on Alberta's boreal forest*", The Pembina Institute and the Canadian Parks and Wilderness Society, Edmonton, AB.

Seigel, Daniel R. (1985) "Estimating Potential Social Losses from Market Failure: Oil Exploration in Alberta." *Rand Journal of Economics*, Vol. 16 (4): 537 – 552.

Severson-Baker, C. 2004 "Exploration and Production Drilling: Environment and Energy in the North." Pembina Institute, Drayton Valley, AB. (Available at www.pembina.org).

Swanson, N.R and White, H. 1997, "Forecasting Economic Time Series Using Flexible Versus Fixed Specification and Linear versus Nonlinear Econometric Models." *International Journal of Forecasting*, (13): 439–461.

Trendle, B. 2006 "Regional Economic Instability, the Role of Industrial Diversification and Spatial Spillovers." *The Annals of Regional Science* Vol.40 (4): 767 – 778.

Uhler, Russel. S. (1979). "The Rate of Petroleum Exploration and Extraction". In *R. S.Pindyck (ed.), Advances in the Economics of Energy and Resources*, Vol. 2. Connecticut, JAI Press Inc., pp. 93–118.

Uhler, Russell. S. 1976, "Costs and Supply in Petroleum Exploration: The Case of Alberta," *Canadian Journal of Economics* 9, 72-90.

US EIA (Energy Information Administration) 2005 http://www.eia.doe.gov.

Vinha K. 2006 "A primer on propensity score matching estimators." Universidad De Los Andes, *CEDE # 001928.*

Walls, Margaret A 1992 "Modeling and Forecasting the Supply of Oil and Gas: A Survey of Existing Approaches" R*esources and Energy,* (14): 287 – 309.

Walls, Margaret A 1994 "Using a 'Hybrid' Approach to Model Oil and Gas Supply: A Case Study of the Gulf of Mexico Outer Continental Shelf" *Land Economics*, 70 (1):1-19.

Weclaw, P. and Hudson, B.J. 2004 "Simulation of conservation and management of woodland caribou. *Ecological Modeling177:75*-94.

Zabel, J.E. and Paterson, R.W. 2006 "The Effects of Critical Habitat Designation on Housing Supply: An Analysis of California Housing Construction Activity." *Journal of Regional Science* 46 (1), 67–95.

Zhang, D. 2004. "Endangered Species and Timber Harvesting: The Case of Red-cockaded Woodpeckers. *Economic Inquiry* 42:150-165.

Appendixes

Appendix 1.1 Alberta Township Survey

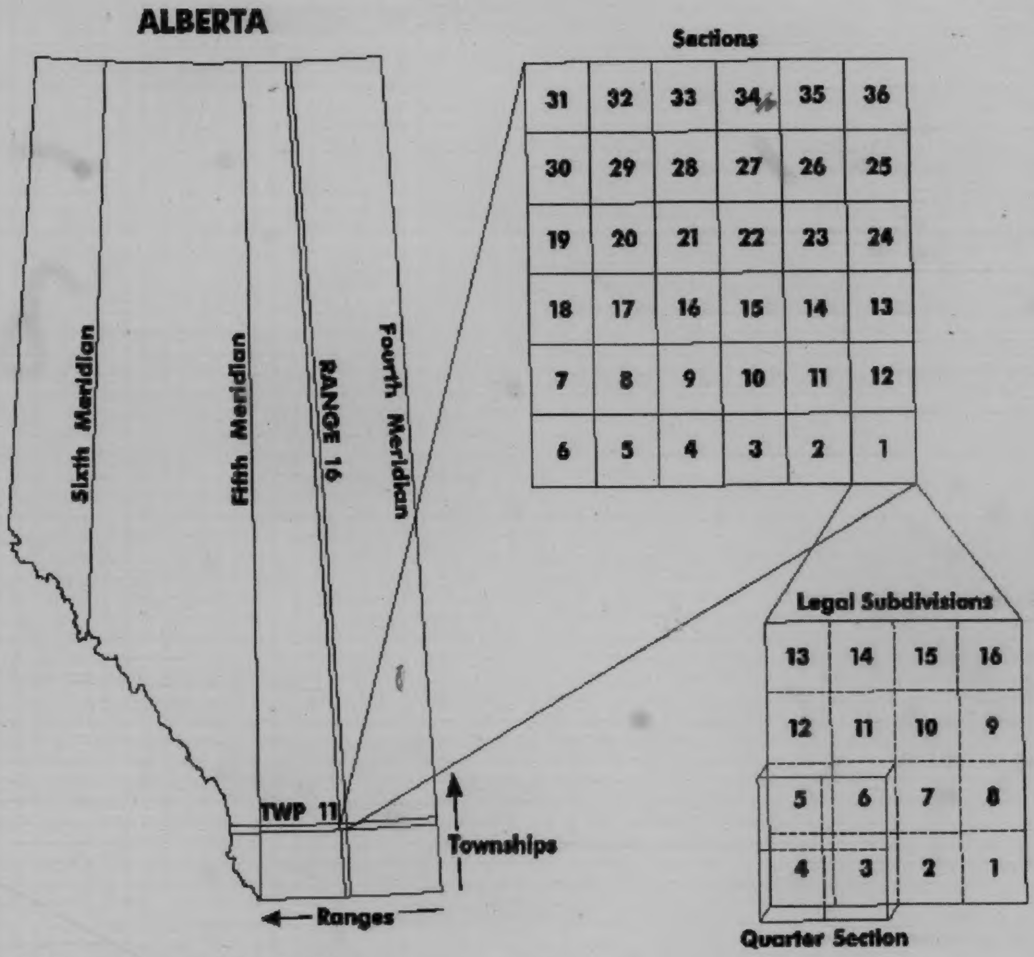
Any parcel of land in Alberta can be located by its legal land description. Legal land descriptions are based on the Alberta township Survey (ATS) system. The ATS is a grid network dividing the province into equal-sized parcels of land.

Under the ATS, land is designated as being west of the 4th, 5th, or 6th Meridians (110°, 114°, 118° west longitude, respectively). Between meridians are six-mile-wide columns called "ranges". Ranges are numbered consecutively from east to west starting at Range 1 west of each meridian. "Townships" are six-mile-wide rows that intersect ranges and are numbered consecutively from Township 1 at the Montana border to Township 126 at the Northwest Territories border.

The word Township also describes the six by six mile square formed by the intersection of ranges and townships. Townships are divided into 36 sections, each section measuring one by one mile. Sections can then be divided into quarters (NE, NW, SE, SW), or into 16 legal subdivisions (LSDs), as indicated.

The legal description of the section highlighted in the diagram would be written as:

1 - 11 - 16 W4
Sec. Twp. Rge. Meridian


ALBERTA
Sixth Meridian
Fifth Meridian
RANGE 16
Fourth Meridian
TWP 11
Townships
Ranges
Sections
31 32 33 34 35 36
30 29 28 27 26 25
19 20 21 22 23 24
18 17 16 15 14 13
7 8 9 10 11 12
6 5 4 3 2 1
Legal Subdivisions
13 14 15 16
12 11 10 9
5 6 7 8
4 3 2 1
Quarter Section

Copyright © 2000, Prairie Conservation Forum

Appendix 2.1 Derivation of optimal exploration equation

A firm decides on the optimal exploration effort (w_{it}) and extraction of oil and gas (q_{it}) by maximizing the expected present value of profits subject to the constraints. The word expected is used to capture the stochastic nature of the optimization problem. Formally this can be written as follows:

$$\underset{w_{it},q_{it}}{Max}.E\int_0^T\left\{\sum_{i=1}^{N}P_t q_{it}-\sum_{i=1}^{N}C^{1i}(w_{it},\Delta_{it})-\sum_{i=1}^{N}C^{2i}(q_{it},R_{it},\Lambda_{it})\right\}e^{-\delta t}dt \qquad 2.1$$

Subject to: $$dR_{it}/dt = d_{it} - q_{it} \text{ and } \quad q_{it} < R_{it} \qquad 2.2$$

$$d_{it} = f(w_{it}, A_{it-1})^{50} \qquad 2.3$$

$$q_{it} \geq 0, w_{it} \geq 0, R_{it} \geq 0, A_{it} \geq 0 \qquad 2.4$$

where equation 2.1 is the objective function, with E as expectation operator and equations 2.2 to 2.4 are constraints. Symbols used in the equations are defined in Table 2.1

Before solving the optimal time path for exploration the following notes are in order. Revenues are generated from the sale of oil and gas extracted at time t from each township i. The price of oil and gas is assumed to be the expected well head price at time t. Producers have two components of costs. The first component is the cost of exploration denoted as $C^1(w_{it}, \Delta_{it})$. This cost increases with drilling effort at a non-

---

[50] In equation 2.3 A_{it-1} refers to lag of cumulative exploratory effort. This specification is used based on the empirical results by Uhler (1976) and Pesaran (1990).

decreasing rate ($\partial C^1(\cdot)/\partial w_{it} > 0$ and $C^1_{ww} \geq 0$). The cost of exploration is also a function of a vector of exogenous physical characteristics such as infrastructure and technological adoptions pertinent to exploration (Δ_{it})[51].

Table 2.1 Definition of the symbol for equations 2.1 – 2.4

Symbol	Description
$t = 0 \ldots T$	Time period from t = 0 to t = T, where T is terminal period.
$i = 1 \ldots N$	Number of grid cells or townships in our case
w_{it}	Drilling effort at township i in period t
q_{it}	Extraction of a resource at township i in period t
P_t	Well head price of a resource at time t
δ	Discount rate
$C^{1i}(.)$	Cost of exploration at township i
$C^{2i}(.)$	Cost of extraction at township i
R_{it}	Proven reserves at township i in period t.
d_{it} , A_{it-1}	Reserve additions and cumulative exploratory effort at township i in period t respectively.
Δ_{it}, Λ_{it}	Vector of exogenous physical characteristics, such as technology, related to exploration and extraction at township i in period t respectively.

The other component of the cost function is cost of development and extraction denoted as: $C^2(q_{it}, R_{it}, \Lambda_{it})$. This cost function is a convex function which increases with the rate of extraction at a non-decreasing rate ($\partial C^2(\cdot)/\partial q_{it} > 0$ and $C^2_{qq} \geq 0$) and decreases with the level of remaining reserves at non-decreasing rate ($\partial C^2(\cdot)/\partial R_{it} < 0$ and $C^2_{RR} \geq 0$)[52]. Λ_{it} captures technological and infrastructure variables in which the cost could be fixed or variable depending on the type of infrastructure and technology pertaining to exploration. In the theoretical model it is assumed that these variables have fixed costs. The inclusion of reserves in the cost function has important implications for firm's extraction and exploration policies. By reducing the level of available reserves current extraction raises future extraction costs, while current exploratory effort tends to lower extraction costs by adding new reserves. Viewed from this perspective exploratory activity can be seen as a way of keeping down marginal extraction costs in the future (Livernois and Uhler 1987).

---

[51] Costs of infrastructure and technology adoption could be variable, however to simplify the analytical solution of the optimization problem, these costs are assumed to be fixed.

In solving equation 2.1 producers face reserve constraints given by equation 2.2. This equation shows that the change in reserves (dR/dt) is explained by the difference between reserve additions and extraction. Current reserve additions are in turn determined by exploratory effort (w_{it}) and lag of cumulative exploratory effort A_{it-1}. Furthermore, firms can not extract more than current reserves (q < R). In equation 2.3, reserve additions are positively related to exploratory effort, $\partial d_{it}(\cdot)/\partial w_{it} = f_w > 0$, and inversely related to lagged cumulative exploratory effort, $\partial d_{it}(\cdot)/\partial A_{it-1} = f_A < 0$. It is worth noting that initially the effect of cumulative exploration on current additions to reserves could be positive due to the influence of accumulated geological knowledge.

---

[52] Livernois and Uhler (1987) argued that although extraction costs tend to rise as reserves are depleted at the intensive margin, these costs also rise as reserves are added at the extensive margin because of the tendency for the least cost deposits to be found first. Hence, the sign of the cost function could be positive or negative depending on the dominance of the cost effect.

However, as exploration continues the effect of reserve exhaustion begins to dominate and beyond a certain threshold value of A_{it-1} the discovery rates start to decline (Quyen 1991).

In addition to the above mentioned constraints, firms consider historical information which include a complex set of geological, economic, and institutional components in their decision making process. For example, historical observations of reserves, price trends, *success rate*s of exploratory effort, technological development, the taxation system, and other government regulations are some of the factors that might affect firms' decision behavior. To simplify the analysis these components are not explicitly included in the theoretical optimization equation.

Equations 2.1 to 2.4 define the decision environment of the firm. The current value Hamiltonian for equation 2.1 and its constraints can be written as:

$$\begin{aligned}\tilde{H} &= E\Big(\sum_{i=1}^{N} P_t q_{it} - \sum_{i=1}^{N} C^{1i}(w_{it}, \Delta_{it}) - \sum_{i=1}^{N} C^{2i}(q_{it}, R_{it}, \Lambda_{it}) \\ &\quad + \lambda_{1it}(d_{it} - q_{it}) + \lambda_{2it} f(w_{it}, A_{it-1})\Big)\end{aligned} \qquad 2.5$$

Replacing $d_{it} = f(w_{it}, A_{it-1})$, equation 2.5 can re-written as:

$$\begin{aligned}\tilde{H} &= E\Big(\sum_{i=1}^{N} P_t q_{it} - \sum_{i=1}^{N} C^{1i}(w_{it}, \Delta_{it}) - \sum_{i=1}^{N} C^{2i}(q_{it}, R_{it}, \Lambda_{it}) \\ &\quad + \lambda_{1it}((f(w_{it}, A_{it-1}) - q_{it})) + \lambda_{2it} f(w_{it}, A_{it-1})\Big)\end{aligned} \qquad 2.6$$

In equation 2.6, R and A are state variables and q and w are control variables. λ_1 and λ_2 are co-state variables and refer to shadow prices of additional units of reserves and cumulative reserve discoveries respectively. The first order conditions for the producer's optimization problem can be obtained by differentiating the current value Hamiltonian with respect to q_{it}, w_{it}, R_{it} , A_{it-1}, λ_1 and, λ_2 as follows. Note: letter subscripts denote partial derivatives and a dot above a letter shows rate of change with respect to time.

$$\partial\tilde{H}/\partial q_{it} = E(p_t - C^{2i}_{q_{it}}(.) - \lambda_{1it}) = 0 \qquad 2.7$$

$$\partial\tilde{H}/\partial w_{it} = E(-C^{1i}_{w_{it}}(.) + f_{w_{it}} \times (\lambda_{1it} + \lambda_{2it})) = 0 \qquad 2.8$$

$$-\partial\tilde{H}\big/\partial R_{it} = E(\dot{\lambda}_{1i} - \delta\lambda_{1it}) = E(C^{2i}_{R_{it}}(.)) \qquad 2.9$$

$$-\partial\tilde{H}\big/\partial A_{it-1} = E(\dot{\lambda}_{2i} - \delta\lambda_{2it}) = E(-f_{A_{it}-1} \times (\lambda_{1it} + \lambda_{2it})) \qquad 2.10$$

$$\partial\tilde{H}/\partial\lambda_{1it} = E(f(w_{it}, A_{it-1}) - q_{it}) = 0 \qquad 2.11$$

$$\partial\tilde{H}/\partial\lambda_{2it} = E(f(w_{it}, A_{it-1})) = 0 \qquad 2.12$$

In addition, at the terminal date T, the following transversality conditions must hold:

$$\tilde{H}(\cdot)\Big|_{t=T} = 0 \qquad 2.13$$

$$\lambda_{1iT} = \lambda_{2iT} = 0 \qquad 2.14$$

An economic interpretation of the above necessary conditions merits discussion. Equation 2.7 shows that at the optimum the expected shadow price of reserves in the ground (λ_1) is given by the difference between the expected well-head price (P_t) and the expected marginal cost of extraction (C_q^2). Equation 3.9 explains the inter-temporal condition for the extraction of oil and gas over time. It states that at the optimum the expected resource rent should be equal to the discounted value of future expected resource rent minus the discounted expected extraction cost due to changes in the reserve base(C_R^2). If we assume that expected costs of extraction are independent of reserves, then equation 3.9 simplifies to the familiar Hotelling rule, which states that at the optimum the expected resource rent should grow at the expected rate of discount.

Equations 2.8 and 2.10 give the necessary conditions for the determination of the optimum level of expected exploratory activity. Equation 2.8 shows that the expected net return to exploration ($f_w\lambda_2$) is the difference between the expected value of exploratory effort ($f_w\lambda_1$) and the expected marginal cost of exploration (C_w^1). Equation 2.10 gives the inter-temporal equilibrium condition for λ_2. It states that at the optimum, the expected net return to exploration should be equal to the discounted value of future net returns to exploration minus the discounted change in the value of expected marginal product of exploration due to the cumulative effect of exploration. Equations 2.11 and 2.12 are derivatives of the Hamiltonian with respect to the co-state variables. Equations 2.13 shows that the value of the Hamiltonian – which measures the total surplus net of the opportunity cost of the resource being depleted – be zero at the terminal date T. Equation 2.14 says that the value of the remaining reserves is zero at the terminal date T.

Since functional forms are not specified for the cost and reserve addition functions, it is not possible to find analytical solutions for the optimal path of exploratory effort (w_{it}). However, optimal time path can be obtained by manipulating the optimality conditions by solving the above equation using optimal control procedures. Solving the above conditions yields:

$$\frac{dw_i}{dt} = E\left(\frac{C^{1i}_{w_{it}}(\cdot)\left[(f_{(wA)_{it}} / f_{w_{it}}) \cdot f(w_{it}, A_{it-1}) - f_{A_{it}} + \delta\right] + C^{2i}_{R_{it}}(\cdot) f_{w_{it}}}{\left[C^{1i}_{(ww)_{it}} - C^{1i}_{w_{it}}(\cdot) f_{(ww)_{it}} / f_{w_{it}})\right]}\right) \qquad 2.5$$

Equation 2.5 shows the time path of exploratory effort. The equation implies that drilling effort is determined by a complex interaction between the expected cost of finding new reserves or expected marginal exploration costs (C^1_w), expected marginal product of exploratory effort (f_w), expected marginal extraction cost due to stock effects (C^2_R), expected reserve additions $f(w, A)$, and the expected rate of interest. Expected costs of exploration and extraction are in turn affected by the level of technology, initial level of reserves and price of the resource. Moreover, expected reserve additions are determined by the level of previous exploration and cumulative exploratory effort.

Appendix 2.2 Matlab routines for estimating spatial panel data models

1. Spatial lag model

```
function results = sar_panel(y,x,W,T,info)
% PURPOSE: computes spatial lag model estimates for spatial panels (N
regions*T time periods)
%           y = p*W*y + X*b + e, using sparse matrix algorithms
% Supply data sorted first by time and then by spatial units, so first
region 1, % region 2, et cetera, in the first year, then region 1,
region 2, etcetera in the second year, and so on
% sem_panel computes y and x in deviation of the spatial and/or time

% (see Baltagi, 2001, Econometric Analysis of Panel Data, ch. 2 and ch.
3)
% ---------------------------------------------------
%  USAGE: results = sar_panel(y,x,W,T,info)
%  where:  y = dependent variable vector
%          x = independent variables matrix
%          W = spatial weights matrix (standardized)
%          T = number of points in time
%  info = an (optional) structure variable with input options:
%  info.model = 0 pooled model without fixed effects (default, x may
      contain an intercept)
%       = 1 spatial fixed effects (x may not contain an intercept)
%       = 2 time period fixed effects (x may not contain an intercept)
%        = 3 spatial and time period fixed effects (x may not contain
      an intercept)
%       info.rmin  = (optional) minimum value of rho to use in search
%       info.rmax  = (optional) maximum value of rho to use in search
%       info.convg = (optional) convergence criterion (default = 1e-8)
%       info.maxit = (optional) maximum # of iterations (default = 500)
%       info.lflag = 0 for full lndet computation (default = 1,fastest)
%        = 1 for MC lndet approximation (fast for very large problems)
%        = 2 for Spline lndet approximation (medium speed)
%info.order = order to use with info.lflag = 1 option (default = 50)
%info.iter  = iterations to use with info.lflag = 1 option (default =
30)
%info.lndet = a matrix returned by sar, sar_g, sarp_g, etc.
%             containing log-determinant information to save time
% ---------------------------------------------------
%  RETURNS: a structure
%         results.meth  = 'psar' if infomodel=0
%                       = 'sarsfe' if info.model=1
%                       = 'sartfe' if info.model=2
%                       = 'sarstfe' if info.model=3
%         results.beta  = bhat
%         results.rho   = rho (p above)
%         results.tstat = asymp t-stat (last entry is rho=spatial
          autoregressive coefficient)
%         results.yhat  = yhat = [inv(y-p*W)]*x*b
%         results.resid = residuals = y-p*W*y-x*b
%         results.sige  = sige = (y-p*W*y-x*b)'*(y-p*W*y-x*b)/n
%         results.rsqr  = rsquared
%         results.rbar  = rbarsquared
%         results.sfe   = spatial fixed effects (if info.model=1 or 3)
%         results.tfe   = time period fixed effects (if info.model=2 or
3)
%         results.con   = intercept (if info.model=3)
%         results.lik   = log likelihood
%         results.nobs  = # of observations
%         results.nvar  = # of explanatory variables in x
%         results.tnvar = nvar + W*y + # fixed effects
%          results.y     = y data vector
%          results.iter  = # of iterations taken
%          results.rmax  = 1/max eigenvalue of W (or rmax if input)
%          results.rmin  = 1/min eigenvalue of W (or rmin if input)
%          results.lflag = lflag from input
%          results.liter = info.iter option from input
%          results.order = info.order option from input
%          results.limit = matrix of [rho lower95,logdet approx, upper95]
           intervals
%                          for the case of lflag = 1
%          results.time1 = time for log determinant calcluation
%          results.time2 = time for eigenvalue calculation
%          results.time3 = time for hessian or information matrix
           calculation
%          results.time4 = time for optimization
%          results.time  = total time taken
%          results.lndet = a matrix containing log-determinant
           information
%                           (for use in later function calls to save
           time)
% --------------------------------------------------
%  NOTES: if you use lflag = 1 or 2, info.rmin will be set = -1
%                                    info.rmax will be set = 1
%         For number of spatial units < 500 you should use lflag = 0 to
get exact results
% --------------------------------------------------
%
% written by: J.Paul Elhorst 11/2004
% University of Groningen
% Department of Economics
% 9700AV Groningen
% the Netherlands
% j.p.elhorst@eco.rug.nl
%
% REFERENCES:
% "Specification and Estimation of Spatial Panel Data Models",
% International Regional Science Review, Vol. 26, pp. 244-268.
% Formulas for information matrix are not in this paper, I derived them
% later.

% This function is based on James. P LeSage's function SAR
```

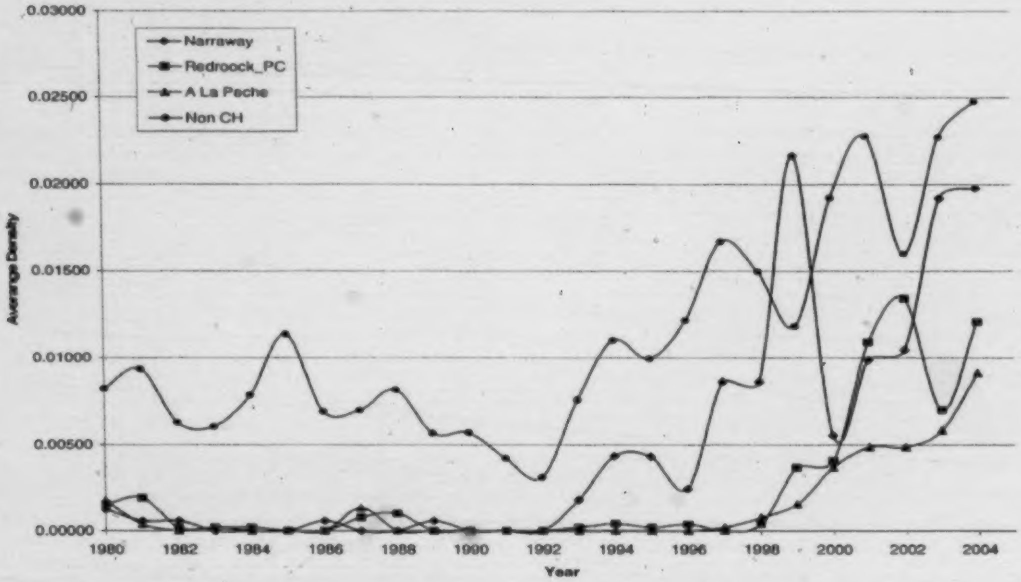
2. Spatial error model

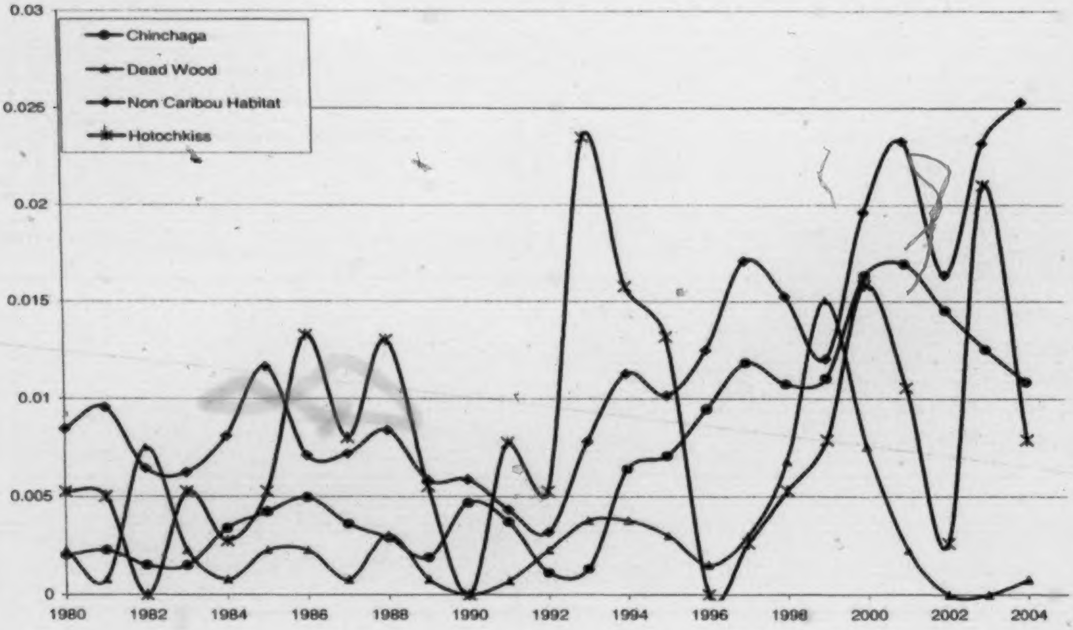
```
function results = sem_panel(y,x,W,T,info)
% PURPOSE: computes spatial error model estimates for spatial panels (N
regions*T time periods)
%           y = XB + u,  u = p*W*u + e, using sparse algorithms
% Supply data sorted first by time and then by spatial units, so first
region 1,
% region 2, et cetera, in the first year, then region 1, region 2, et
% cetera in the second year, and so on
% sem_panel computes y and x in deviation of the spatial and/or time
means
% (see Baltagi, 2001, Econometric Analysis of Panel Data, ch. 2 and ch.
3)
% --------------------------------------------------
%  USAGE: results = sem_panel(y,x,W,T,info)
%  where: y = dependent variable vector
%         x = independent variables matrix
%         W = spatial weights matrix (standardized)
%         T = number of points in time
%       info = an (optional) structure variable with input options:
```

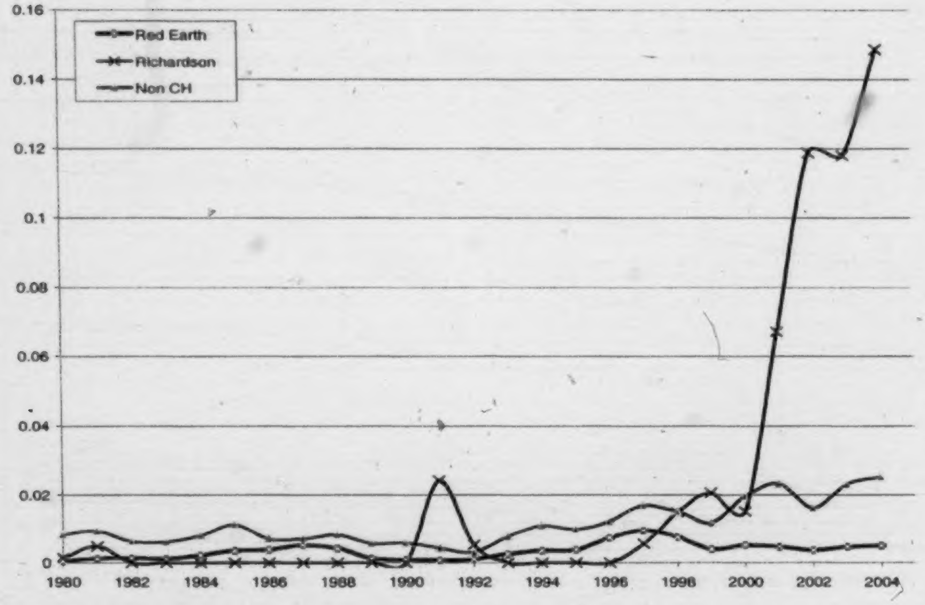
```
%        info.model = 0 pooled model without fixed effects (default, x
        may contain an intercept)
%                   = 1 spatial fixed effects (x may not contain an
        intercept)
%                   = 2 time period fixed effects (x may not contain an
        intercept)
%                   = 3 spatial and time period fixed effects (x may not
        contain an intercept)
%        info.rmin  = (optional) minimum value of rho to use in search
%        info.rmax  = (optional) maximum value of rho to use in search
%        info.convg = (optional) convergence criterion (default = 1e-4)
%        info.maxit = (optional) maximum # of iterations (default = 500)
%        info.lflag = 0 for full lndet computation (default = 1, fastest)
%                   = 1 for MC lndet approximation (fast for very large
        problems)
%                   = 2 for Spline lndet approximation (medium speed)
%        info.order = order to use with info.lflag = 1 option (default =
50)
%        info.iter  = iterations to use with info.lflag = 1 option
         (default = 30)
%        info.lndet = a matrix returned by sar, sar_g, sarp_g, etc.
%                     containing log-determinant information to save time
% ---------------------------------------------------------
% RETURNS: a structure
%          results.meth  = 'psem' if infomodel=0
%                        = 'semsfe' if info.model=1
%                        = 'semtfe' if info.model=2
%                        = 'semstfe' if info.model=3
%          results.beta  = bhat
%          results.rho   = rho (p above)
%          results.tstat = asymp t-stats (last entry is rho=spatial
           autocorrelation coefficient)
%          results.yhat  = yhat
%          results.resid = residuals
%          results.sige  = sige = e'(I-p*W)'*(I-p*W)*e/nobs
%          results.rsqr  = rsquared
%          results.rbar  = rbarsquared
%          results.sfe   = spatial fixed effects (if info.model=1 or 3)
%          results.tfe   = time period fixed effects (if info.model=2 or
3)
%          results.con   = intercept (if info.model=3)
%          results.lik   = log likelihood
%          results.nobs  = # of observations
%          results.nvar  = # of explanatory variables in x
%          results.tnvar = nvar + # fixed effects
%          results.y     = y data vector
%          results.iter  = # of iterations taken
%          results.rmax  = 1/max eigenvalue of W (or rmax if input)
%          results.rmin  = 1/min eigenvalue of W (or rmin if input)
%          results.lflag = lflag from input
%          results.liter = info.iter option from input
%          results.order = info.order option from input
%          results.limit = matrix of [rho lower95,logdet approx, upper95]
           intervals
%                          for the case of lflag = 1
%          results.time1 = time for log determinant calcluation
%          results.time2 = time for eigenvalue calculation
%          results.time3 = time for hessian or information matrix
           calculation
%          results.time4 = time for optimization
%          results.time  = total time taken
%          results.lndet = a matrix containing log-determinant
           information
```

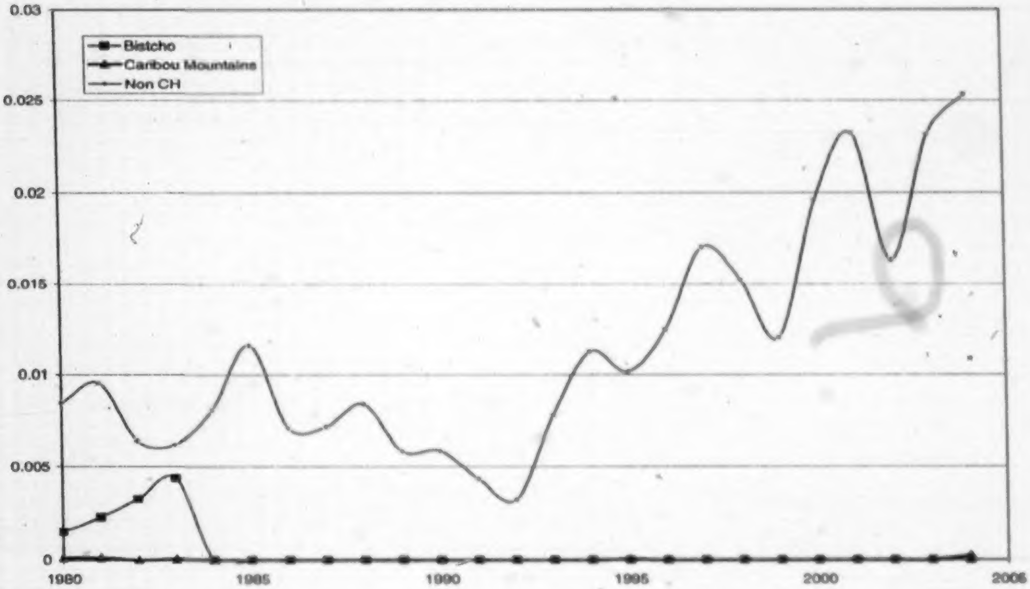
```
%                    (for use in later function calls to save
         time)
% ------------------------------------------------
% NOTES: if you use lflag = 1 or 2, info.rmin will be set = -1
%                                   info.rmax will be set = 1
%        For number of spatial units < 500 you should use lflag = 0 to
get exact results
% ------------------------------------------------
%
% written by: J.Paul Elhorst 11/2004
% University of Groningen
% Department of Economics
% 9700AV Groningen
% the Netherlands
% j.p.elhorst@eco.rug.nl
%
% REFERENCES:
% "Specification and Estimation of Spatial Panel Data Models",
% International Regional Science Review, Vol. 26, pp. 244-268.
```

Appendix 6.1 Average drilling densities by herd name

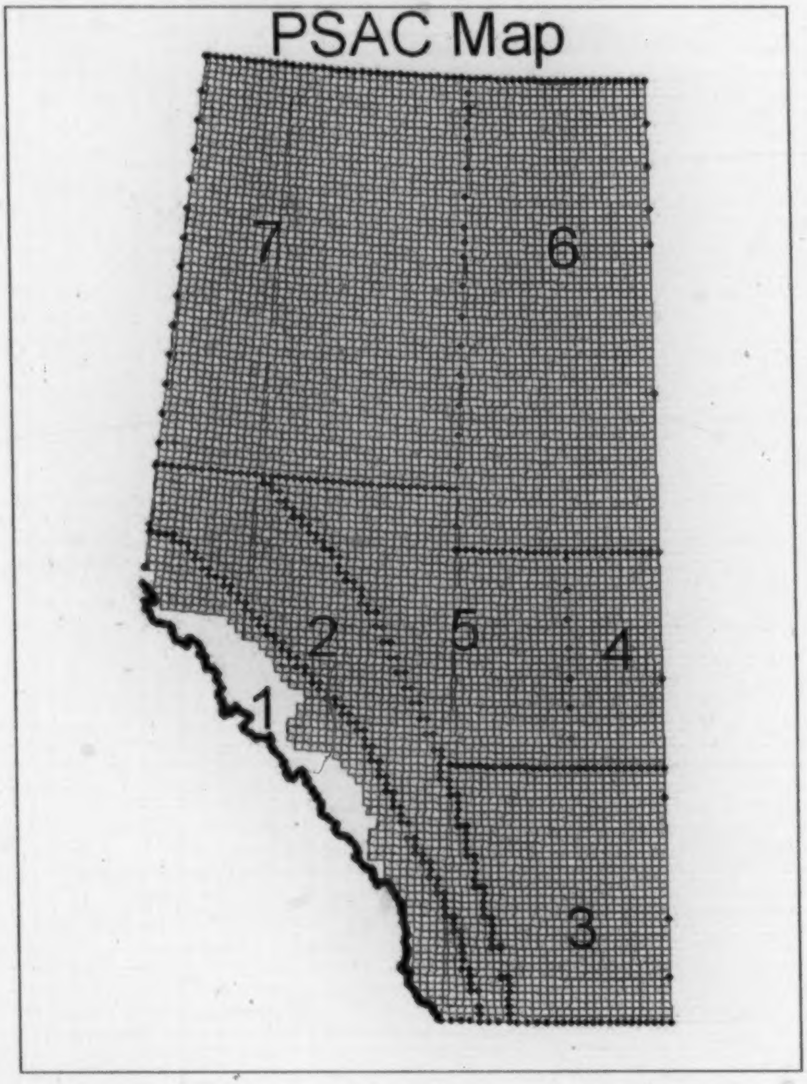



Chinchaga
Dead Wood
Non Caribou Habitat
Hotochkiss
0.03
0.025
0.02
0.015
0.01
0.005
0
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004


Red Earth
Richardson
Non CH
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004


0.03
0.025
0.02
0.015
0.01
0.005
0
1980
1985
1990
1995
2000
2005
Bistcho
Caribou Mountains
Non CH

Appendix 6.2 Map based on Petroleum Services Association of Canada (PSAC)


PSAC Map
7
6
2
5
4
1
3

Source: PSAC.

Appendix 6.3 Results for the herd by herd multivariate regression model

Variable	Coefficient	Standard Error	T-stat
Cummulative wells	5.50E-02	9.39E-04	58.544
Cum. Wells square	-2.57E-03	2.12E-04	-12.118
Trend	9.23E-05	1.01E-03	0.092
Lagged success rate	1.82E-02	6.94E-04	26.174
Lagged price	4.55E-03	9.73E-04	4.673
Capacity	1.42E-02	4.67E-03	3.049
Cold Lake	1.87E-03	3.54E-03	0.527
ESAR	3.28E-03	2.46E-03	1.335
WSAR	4.66E-04	2.15E-03	0.217
Red Earth	2.77E-03	2.02E-03	1.375
Richardson	1.29E-02	4.84E-03	2.666
Slave Lake	1.88E-04	5.03E-03	0.037
Chinchaga	2.87E-03	2.62E-03	1.096
Narraway	3.33E-03	6.17E-03	0.54
Hotchick	-5.31E-03	1.24E-02	-0.43
Deadwood	1.80E-03	6.63E-03	0.272
A La Peche	3.91E-03	3.47E-03	1.127
Little Smoky	1.40E-03	4.49E-03	0.312
Red Creek	3.71E-03	3.55E-03	1.044
Caribou Mountain	4.83E-03	1.98E-03	2.434
Bistcho	3.38E-03	2.23E-03	1.515
Time 1993_95 (1)	7.71E-04	1.84E-03	0.419
Time 1996_98 (2)	2.27E-03	2.17E-03	1.044
Time 1999_01 (3)	-6.87E-04	2.13E-03	-0.323
Time 2000_04 (4)	-8.14E-04	2.35E-03	-0.346
*Cold Lake 1**	4.33E-03	5.02E-03	0.862
Cold Lake 2	2.60E-02	5.01E-03	5.198
Cold Lake 3	1.42E-02	5.03E-03	2.823
Cold Lake 4	2.04E-02	5.02E-03	4.065
ESAR 1	5.24E-03	3.48E-03	1.504
ESAR 2	-4.11E-03	3.48E-03	-1.181
ESAR 3	-7.27E-03	3.48E-03	-2.09
ESAR 4	-1.18E-02	3.48E-03	-3.398
WSAR 1	-8.40E-04	3.03E-03	-0.277
WSAR 2	9.18E-03	3.04E-03	3.024
WSAR 3	9.29E-03	3.04E-03	3.057
WSAR 4	3.12E-03	3.03E-03	1.029
Red Earth 1	-1.93E-03	2.85E-03	-0.677
Red Earth 2	-8.29E-04	2.84E-03	-0.291
Red Earth 3	-4.73E-03	2.85E-03	-1.66
Red Earth 4	-5.52E-03	2.85E-03	-1.938
Richardson 1	-1.47E-02	6.83E-03	-2.146
Richardson 2	-9.69E-03	6.83E-03	-1.418
Richardson 3	1.26E-02	6.84E-03	1.844
Richardson 4	9.13E-02	6.82E-03	13.39

Slave Lake 1	2.65E-03	7.12E-03	0.372
Slave Lake 2	-3.85E-03	7.12E-03	-0.541
Slave Lake 3	-1.30E-02	7.11E-03	-1.826
Slave Lake 4	-6.40E-03	7.12E-03	-0.899
Chinchaga 1	-1.71E-03	3.72E-03	-0.46
Chinchaga 2	-1.43E-03	3.71E-03	-0.386
Chinchaga 3	8.58E-04	3.71E-03	0.231
Chinchaga 4	-3.20E-03	3.69E-03	-0.866
Narraway 1	3.03E-04	8.74E-03	0.035
Narraway 2	1.37E-05	8.74E-03	0.002
Narraway 3	4.31E-03	8.74E-03	0.493
Narraway 4	3.92E-03	8.74E-03	0.448
Hotchick 1	-2.92E-04	1.75E-02	-0.017
Hotchick 2	-8.09E-03	1.75E-02	-0.463
Hotchick 3	-8.75E-03	1.75E-02	-0.501
Hotchick 4	-9.18E-03	1.75E-02	-0.525
Deadwood 1	-2.38E-03	9.37E-03	-0.254
Deadwood 2	-3.61E-03	9.37E-03	-0.385
Deadwood 3	-3.99E-03	9.37E-03	-0.426
Deadwood 4	-8.05E-03	9.37E-03	-0.858
A La Peche 1	-3.73E-03	4.93E-03	-0.757
A La Peche 2	-5.59E-03	4.91E-03	-1.137
A La Peche 3	-3.95E-03	4.93E-03	-0.801
A La Peche 4	-3.58E-03	4.91E-03	-0.728
Little Smoky 1	-1.49E-03	6.36E-03	-0.234
Little Smoky 2	-4.58E-03	6.35E-03	-0.72
Little Smoky 3	-2.45E-03	6.36E-03	-0.386
Little Smoky 4	1.19E-02	6.35E-03	1.88
Red Creek 1	-4.61E-03	5.01E-03	-0.919
Red Creek 2	-5.44E-03	5.01E-03	-1.088
Red Creek 3	-6.70E-04	5.02E-03	-0.133
Red Creek 4	-1.47E-03	5.02E-03	-0.293
Caribou Mountain 1	-3.59E-03	2.80E-03	-1.283
Caribou Mountain 2	-6.93E-03	2.80E-03	-2.476
Caribou Mountain 3	-7.11E-03	2.80E-03	-2.536
Caribou Mountain 4	-7.48E-03	2.80E-03	-2.671
Bistcho 1	-3.78E-03	3.15E-03	-1.199
Bistcho 2	-5.58E-03	3.15E-03	-1.769
Bistcho 3	-7.62E-03	3.15E-03	-2.417
Bistcho 4	-9.72E-03	3.15E-03	-3.08
Constant	-1.32E-02	1.33E-03	-9.908

*** Note: Herd name followed by numbers 1 to 4 show interaction terms between each herd and time dummy.**